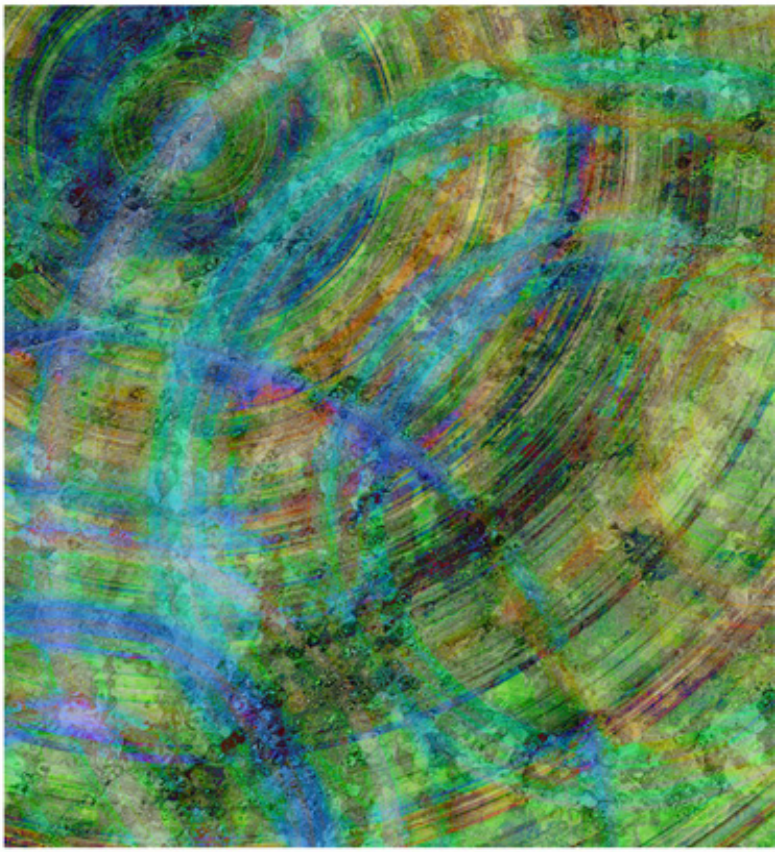
The Handbook of English Pronunciation



Edited by
Marnie Reed and John M. Levis

WILEY Blackwell

The Handbook of English Pronunciation

Blackwell Handbooks in Linguistics

This outstanding multivolume series covers all the major subdisciplines within linguistics today and, when complete, will offer a comprehensive survey of linguistics as a whole.

Recent Titles Include:

The Handbook of Computational Linguistics and Natural Language Processing
Edited by Alexander Clark, Chris Fox, and Shalom Lappin

The Handbook of Language and Globalization
Edited by Nikolas Coupland

The Handbook of Hispanic Sociolinguistics
Edited by Manuel Díaz-Campos

The Handbook of Language Socialization
Edited by Alessandro Duranti, Elinor Ochs, and Bambi B. Schieffelin

The Handbook of Intercultural Discourse and Communication
Edited by Christina Bratt Paulston, Scott F. Kiesling, and Elizabeth S. Rangel

The Handbook of Historical Sociolinguistics
Edited by Juan Manuel Hernández-Campoy and Juan Camilo Conde-Silvestre

The Handbook of Hispanic Linguistics
Edited by José Ignacio Hualde, Antxon Olarrea, and Erin O'Rourke

The Handbook of Conversation Analysis
Edited by Jack Sidnell and Tanya Stivers

The Handbook of English for Specific Purposes
Edited by Brian Paltridge and Sue Starfield

The Handbook of Spanish Second Language Acquisition
Edited by Kimberly L. Geeslin

The Handbook of Chinese Linguistics
Edited by C.-T. James Huang, Y.-H. Audrey Li, and Andrew Simpson

The Handbook of Language Emergence
Edited by Brian MacWhinney and William O'Grady

The Handbook of Korean Linguistics
Edited by Lucien Brown and Jaehoon Yeon

The Handbook of Speech Production
Edited Melissa A. Redford

The Handbook of Contemporary Semantic Theory, Second Edition
Edited by Shalom Lappin and Chris Fox

The Handbook of Classroom Discourse and Interaction
Edited by Numa Markee

The Handbook of Narrative Analysis
Edited by Anna De Fina & Alexandra Georgakopoulou

The Handbook of English Pronounciation
Edited by Marnie Reed and John M. Levis

Full series title list available at www.blackwellreference.com

The Handbook of English Pronunciation

Edited by

Marnie Reed and John M. Levis

WILEY Blackwell

This edition first published 2015

Registered Office
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

Editorial Offices
350 Main Street, Malden, MA 02148-5020, USA
9600 Garsington Road, Oxford, OX4 2DQ, UK
The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

For details of our global editorial offices, for customer services, and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com/wiley-blackwell.

Library of Congress Cataloging-in-Publication Data applied for

9781118314470 (hardback)

A catalogue record for this book is available from the British Library.

Cover image: © Rolffimages | Dreamstime.com

Set in 10/12pt Palatino by SPi Publisher Services, Pondicherry, India

1 2015

Contents

Notes on Contributors

Ghinwa Alameen, PhD, teaches TESL and Arabic at Iowa State University. Her research focuses on the effectiveness of teaching connected speech on L2 perception and production. She has published articles on L2 material design, the integration of technology in language teaching, and the teaching of oral skills.

Amanda A. Baker, PhD, is Coordinator of the TESOL program at the University of Wollongong in Australia. Amanda's research interests focus on the dynamic relationships that exist between second language (L2) teachers' knowledge, beliefs, and practices, especially in the areas of L2 pronunciation, speaking, and listening pedagogy.

Laurie Bauer is Emeritus Professor of Linguistics at Victoria University of Wellington, New Zealand. He has published widely on international varieties of English, especially New Zealand English, and on morphology. Most recently, he is one of the authors of the *Oxford Reference Guide to English Morphology* (2013).

Ian Bekker, currently at the Potchefstroom campus of the North-West University, specializes in the sociophonetics of South African English (SAfE), both in terms of contemporary developments as well as the reconstruction of its past genesis and development. His current main research focus is on the role of Johannesburg in the development of SAfE.

Charles Boberg teaches Linguistics at McGill University in Montreal, Canada. His research focuses on variation and change in North American English, particularly Canadian English. His books include *The English Language in Canada: Status, History and Comparative Analysis* (2010) and, with William Labov and Sharon Ash, *The Atlas of North American English* (2006).

Adam Brown is the Director of Research at Auckland Institute of Studies, New Zealand. He holds a PhD in phonetics from the University of Edinburgh and has taught in Thailand, Malaysia, and Singapore, as well as his native UK. He has written a dozen books and many articles on aspects of English language teaching, especially pronunciation. His latest publication is *Pronunciation and Phonetics: A Practical Guide for English Language Teachers* (2014).

Graeme Couper is a senior lecturer at Auckland University of Technology with many years teaching experience in a wide range of countries and contexts, which he applies to his research into the teaching and learning of L2 pronunciation. His classroom-based research brings theory and practice together, finding a significant role for Cognitive Linguistics and other usage-based theories that allow for both the cognitive and social nature of language learning.

Anne Cutler is Emeritus Director of the Max Planck Institute for Psycholinguistics, Professor at the MARCS Institute, University of Western Sydney, and Processing Program leader of the Australian Research Council Centre of Excellence in the Dynamics of Language. Her research (summarized in her 2012 book *Native Listening*) focuses on how native-language phonological structure shapes the way we listen to speech.

Isabelle Darcy is Associate Professor of second language psycholinguistics in the Department of Second Language Studies at Indiana University. She obtained a PhD in Linguistics and Cognitive Science from the Ecole des Hautes Etudes en Sciences Sociales in Paris (France) and from the Johannes Gutenberg University in Mainz (Germany); her research focuses on the acquisition of second language phonology, pronunciation instruction, native/non-native speech perception, and word recognition.

Tracey M. Derwing is a Professor Emeritus in TESL (Department of Educational Psychology) at the University of Alberta and an Adjunct Professor in the Department of Linguistics at Simon Fraser University. Her research interests include L2 pronunciation, native speaker reactions to accented speech, pragmatics, immigration, settlement, and teacher education. Together with Murray Munro, she conducted a 10-year longitudinal study of naturalistic pronunciation development in two groups of language learners.

David Deterding is a Professor at Universiti Brunei Darussalam, where he teaches phonetics, grammar, research methods, translation, and forensic linguistics. His research focuses on acoustic phonetics, the pronunciation of Chinese and Malay, the description of Englishes in Southeast Asia, and misunderstandings in English as a lingua franca.

Wayne B. Dickerson is Professor Emeritus in the Department of Linguistics at the University of Illinois at Urbana-Champaign where he taught courses in English phonology for MATESL candidates and ESL pronunciation. His research focuses on pedagogical applications of phonetics and phonology, pronunciation pedagogy, the value of orthography for learners, phonological variability, and pronunciation assessment.

Jennifer Ann Foote is a doctoral student at Concordia University, Montreal, Canada. She has taught English in Canada, Japan, the Czech Republic, and South Korea. She is interested in issues related to teaching pronunciation.

Rebecca Hincks is Associate Professor of English at the KTH Royal Institute of Technology in Stockholm, Sweden. Her doctoral work was done in language-learning

applications for speech processing. Her research interests are oriented toward the development of training systems for public speaking in a lingua franca environment.

Seong-Yoon Kang, PhD, is an Associate Director of International Teacher and Government Programs and Curriculum Specialist at Bloomfield College, USA, where he is in charge of Total Immersion Courses for Korean English Teachers (TICKET) as well as intensive English programs. His research focuses on L2 learners' individual differences in language acquisition and sociolinguistic influences on speech acts. Previously he designed, developed, and taught intensive English courses in South Korea.

Sara Kennedy is an Associate Professor of Applied Linguistics in the Department of Education at Concordia University, Montreal, Canada. Her research focuses on intelligibility of second language speech, effects of classroom instruction, particularly the teaching of oral skills, and the role of language experience in the development of speaking ability. She has extensive experience teaching English as a second and foreign language.

John M. Levis is Professor of Applied Linguistics and TESL at Iowa State University. His research interests are English intonation, teacher education for pronunciation, and speech intelligibility. He is the editor of the *Journal of Second Language Pronunciation*.

Ee-Ling Low PhD (Cambridge, UK) is an Associate Professor of English Language and Literature and concurrently Head of Strategic Planning and Academic Quality at the National Institute of Education, Singapore. She has published widely in the areas of the phonetics of World Englishes and pronunciation for English as an International Language.

Christina Michaud is a Senior Lecturer in the Writing Program at Boston University, where she teaches argument and research writing to native and non-native speakers of English. She has co-authored a supra-segmental pronunciation textbook and a book on lesson planning for TESOL teachers.

Lynda Mugglestone is Professor of History of English at the University of Oxford and a Fellow of Pembroke College, Oxford. She has published widely on the history of English, with special interests in the history of pronunciation, and in metalexicography and the social, cultural, and ideological issues that dictionary-making can reveal.

Murray J. Munro, a Professor of Linguistics at Simon Fraser University, has published extensively on accent and intelligibility, vowel and consonant acquisition, and the role of age and experience in phonetic learning. His collaborative work with Tracey Derwing focuses on the empirical foundations of pronunciation teaching.

John M. Murphy is a Professor of Applied Linguistics and ESL at Georgia State University (Atlanta). His research and pedagogic interests span three areas: second

language (L2) teacher reasoning (e.g., teachers' knowledge, beliefs, and instructional practices), approaches to L2 teaching, and integrated instruction of ESL listening, speaking, and pronunciation. John also teaches Yoga (twice weekly) in the lineage of Pranakriya Hatha Yoga.

Cecil L. Nelson is the author of *Intelligibility in World Englishes* (2011) and a co-editor with Braj and Yamuna Kachru of *The Handbook of World Englishes* (2006). He was for some years the Review Editor of the journal *World Englishes*.

Pramod Pandey is Professor of Linguistics, Centre for Linguistics at Jawaharlal Nehru University, New Delhi. His current areas of research interest include phonological interfaces, Indian English, writing systems, speech technology, and multilingualism. His publications include research articles on phonetics-phonology, second language varieties, writing systems, and a recent book, *Sounds and Their Patterns in Indic Languages* (two volumes).

Marnie Reed is an Associate Professor of Education and affiliated faculty in Applied Linguistics at Boston University. Her research focuses on second language phonology, particularly the role of auditory feedback in the perception and production of connected discourse, the role of metalinguistic feedback in the acquisition of morphosyntax, and metacognition in cross-linguistics awareness of pragmatic functions of English prosody.

Laura Sicola PhD is a lecturer in the MS-TESOL program at the University of Pennsylvania's Graduate School of Education, where she received her PhD in educational linguistics. Her primary research is in L2 pronunciation pedagogy and the use of pedagogic tasks. Her company, the Sicola Consulting Group, specializes in business-English communication programs for non-native speakers in professional and executive roles.

Jeremy Smith is Professor of English Philology at the University of Glasgow. His publications include: *Older Scots: A Linguistic Reader* (2012); *Old English: A Linguistic Introduction* (2009); *Sound Change and the History of English* (2007); *An Introduction to Middle English* (with Simon Horobin, 2002); *Essentials of Early English* (second edition, 2005); *An Historical Study of English: Function, Form and Change* (1996). He is currently working on the application of notions from historical pragmatics and book history to the study of medieval and early modern English and Scottish writing systems.

Beatrice Szczepek Reed is Senior Lecturer in Second Language Education at the University of York and her interest is in the phonetics and prosody of natural talk. She publishes in leading peer-reviewed journals and has written *Analysing Conversation: An Introduction to Prosody* (2010) and *Prosodic Orientation in English Conversation* (2006).

Ron I. Thomson is an Associate Professor of Applied Linguistics at Brock University. His research interests include L2 oral fluency and pronunciation, and listener evaluation of L2 speech. He is also the creator of a free, evidence-based English pronunciation training website and research tool – www.englishaccentcoach.com.

Pavel Trofimovich is an Associate Professor of Applied Linguistics in the Department of Education at Concordia University, Montreal, Canada. His research focuses on cognitive aspects of second language processing, second language phonology, sociolinguistic aspects of second language acquisition, and the teaching of second language pronunciation.

Clive Upton is Emeritus Professor of Modern English Language at the University of Leeds, and edits *English Today*. He co-edited *A Handbook of Varieties of English*, co-authored *The Oxford Dictionary of Pronunciation for Current English*, and his transcription system for Received Pronunciation is followed in the *Oxford English Dictionary*.

Bertus van Rooy is Professor of English language studies at the Vaal Triangle Campus of the North-West University, and a past president of the *International Association for World Englishes*. His current research is focused on the features of varieties of English, with particular attention to the development of new features.

Marilyn May Vihman is a developmental linguist best known for her book, *Phonological Development* (1996), which appeared in a radically revised second edition in 2014, with updated surveys of research on infant speech perception, segmentation, distributional learning, experimental studies of word learning, and other aspects of phonological development.

Robin Walker is a freelance teacher, trainer, and materials writer. A long-standing committee member of the IATEFL Pronunciation Special Interest Group, and a former editor of *Speak Out*!, the PronSIG journal, he regularly gives talks, workshops, and webinars on pronunciation teaching. He is the author of *Teaching the Pronunciation of English as a Lingua Franca*, a handbook for teachers.

Anne Wichmann is Emeritus Professor of Speech and Language at the University of Central Lancashire, Preston, UK. Her research focuses on prosody, especially intonation, as it is used to structure discourse and in its role in the expression of pragmatic meaning, including attitude and emotion.

Beth Zielinski is an Honorary Associate in the Department of Linguistics at Macquarie University, Australia. Her research interests are in the area of L2 English speaking, particularly the influence of different pronunciation features on effective communication. Her research and teaching has involved learners of English in many different settings in Australia.

Wafa Zoghbor is an applied linguist and she is currently an Assistant Professor at Zayed University, UAE. Her doctoral thesis in 2010 at the University of Leicester was on the implications of the pronunciation of ELF for Arab learners. In addition to teaching and research, her professional background involves quality assurance in higher education.

Introduction

The Handbook of English Pronunciation is a collection of 28 chapters with various approaches to English pronunciation. As we have worked on the Handbook, we have been strongly aware that we could have doubled the number of chapters and still not fully done justice to the overall topic. The Handbook is intended for applied linguists and for teachers, for those who are experts and for those who are not. In applied linguistics, a growing number of researchers are examining pronunciation and its relationship to areas such as speech intelligibility, language testing, speech recognition and text-to-speech, pragmatics, and social factors impacting language acquisition. Indeed, researchers in any area of applied linguistics increasingly find the need to take phonetic and phonological form into account. They may not be experts in pronunciation, yet still they find a need to understand the forms and meanings of English pronunciation and they need to know where to find further information when they need it. Beyond directly practical chapters, many authors of more research-oriented chapters have added implications of research for teaching.

The handbook is also written for teachers who need immediately practical chapters about the place of pronunciation in their classrooms. They also need a wider context for how English pronunciation is structured, why it is so varied, and how it changes depending on discourse context. This means that the handbook includes chapters that are important in understanding the role of pronunciation in language description and analysis, and chapters that are more obviously relevant to teachers. A single book that tries to meet the needs of both groups is a challenge, but it is also necessary for a field with growing interest both for the classroom and for research.

The handbook is necessary because pronunciation is a topic that will not go away. Pronunciation influences all research into, and teaching of, spoken language, which must take account of how English is pronounced to account for what happens elsewhere in spoken language. Discourse analysis, pragmatics, sociocultural analyses of language, English as an international language, reading, acquisition, and ultimate attainment, all must reckon with pronunciation as an important variable. Those primarily interested in other areas may not be experts in pronunciation, yet still find a need to understand the forms and meanings of

English pronunciation and where to find further information when they need it. Not only is pronunciation important in relation to other areas of language but it is important in its own right.

A knowledge of English pronunciation is also valuable by itself as an area of study. Even though a native-like accent is impossible for most adult L2 learners, pronunciation remains the gateway to spoken intelligibility for second language learners because of its close ties to social meanings within language. It also helps distinguish dialects, formal and informal registers of speech, and is influential in distinguishing social standing within speech networks.

In English language teaching, pronunciation is today on the ascendancy. As a subject area for language teaching, it plummeted from being central to falling into disfavor in the 1960s and 1970s when research confronted teachers with the uncomfortable fact that it was impossible, or at least extraordinarily unlikely, for second language learners to achieve a native-like accent. Additionally, the rise of communicative language teaching and its emphasis on fluency was a poor fit for the 1960s accuracy-oriented exercises of pronunciation teaching. As a result, pronunciation was often ignored in the classroom, with the hope that it would somehow take care of itself if teachers worked on helping learners achieve communicative competence.

Unfortunately, this hope was overly optimistic. Pronunciation did not take care of itself. The two choices of "we need to have native-like pronunciation" versus "it's not worth working on this if we can't be native" have been increasingly shown by research and practice to be a false dichotomy. Hinofotis and Bailey (1981) were among the first to argue that pronunciation played a kind of gate-keeping function in speech, in that speakers who had not achieved a threshold level of pronunciation adequacy in the second language would not, and could not, be adequate communicators no matter how good their fluency, listening, grammar, and vocabulary. The resurrection of the notion of intelligibility (Abercrombie 1949) as both a more reasonable and more realistic goal for pronunciation achievement began with Smith and Nelson's (1985) examination of intelligibility among World Englishes. Their classificatory scheme of intelligibility was mirrored in many ways by research done by James Flege, and Murray Munro and Tracey Derwing (1995) and has had a tremendous effect not only on research into pronunciation learning but also in the way it is approached in the classroom (see Levis 2005).

Even though teachers throughout the world recognize the importance of pronunciation, they have repeatedly reported feeling inadequate in addressing this area of language teaching (Burgess and Spencer 2000; Breitkreutz, Derwing, and Rossiter 2002; Macdonald 2002). As a result of their confusion and lack of confidence, most simply do not address pronunciation. While a full solution to this lack of confidence would require many changes in professional preparation both for teachers and applied linguistics researchers, a reliable, easily available source of information that reflects current knowledge of the field is one important step.

Throughout this Handbook, we learn how an understanding of English pronunciation is essential for any applied linguist or language teacher, from understanding the historical and often unusual development of English pronunciation over 1000 years, to descriptions of the diversity of Englishes and their

pronunciations in the world today, to the ways that features of English pronunciation are best described, to pronunciation's role in the construction and the analysis of discourse, to patterns of first and second language acquisition, and to the social attitudes connected to differences in accent. Even this wide range of topics is too narrow. English pronunciation carries social meanings and is subject to social judgments, it reflects pragmatic meanings, it is intimately connected to the expression of information structure, and it is essential to speech recognition and text-to-speech technology. Pronunciation cannot be ignored.

The structure of the Handbook includes six general areas: History, Description, Discourse, Varieties, Acquisition, and Teaching. The first area tells us of the history of English pronunciation. English has a very interesting history of its pronunciation, going back more than 1000 years. Jeremy Smith provides a long view of how English has changed, looking at residualisms in varieties of English and focusing especially on three major changes: the phonemicization of voiced fricatives, the effect of Breaking on vowel changes, and the Great Vowel Shift. Each of these remains important in today's Englishes, showing that history is not just the past but influences today's Englishes as well. In the second chapter in this section, Lynda Mugglestone examines the social meanings of accent from the eighteenth century until today. The rise of Received Pronunciation (RP) as a marker of education and class both included and excluded speakers from the social power structure and reinforced social class barriers as RP spread throughout the power structure of Great Britain. The chapter is a fascinating look at how important "talking proper" (Mugglestone 2007) was and how even now the values associated with accent remain powerful. Finally, John Murphy and Amanda Baker look at the history of pronunciation teaching from 1850 till now. They identify four overlapping waves of practice, with a fifth wave perhaps in its early stages. Their meticulously researched history of pronunciation teaching will provide a framework for researchers and will help teachers understand where pedagogical approaches originated.

The second section of the Handbook is the bread and butter of pronunciation, the description of the structural units that make up the widely varying elements of the system. David Deterding provides a look at the segmentals of English, focusing his attention on the consonant and vowel sounds. Adam Brown looks at what happens to those segmentals when they are combined into syllables and how certain patterns are well formed and others are not. His discussion of phonotactics is important for anyone looking at acquisition since well-formed structures in English syllables are not always well formed in other languages. Anne Cutler looks at the ever-important but often misunderstood topic of lexical stress. An expert in how English speakers perceive stress and the signals they attend to, Cutler argues that the prosodic and segmental features of lexical stress are redundant and that listeners primarily attend to segmental cues. Ee Ling Low describes English rhythm from a cross-variety standpoint. She looks at how assumptions of stress-timed rhythm are and are not justified and what recent research on rhythmic variation in different varieties of world Englishes tells us about English rhythm and its place in pronunciation teaching. John M. Levis and Anne Wichmann look at the significant uses of pitch to communicate meaning in their chapter on

intonation. Intonation in English is one of the oldest topics to be addressed from an applied viewpoint, yet it remains one of the topics where the gap between modern linguistic descriptions and applied linguistic work is widest. Levis and Wichmann describe newer approaches and the ways in which intonation communicates meaning.

The next section looks at research into how pronunciation behaves at the discourse level. Most research still is done at the sound, word, and sentence level, but discourse affects pronunciation in special ways that are important for both researchers and teachers. Ghinwa Alameen and John M. Levis provide an overview of a much-neglected topic in research, Connected Speech Processes. Comprised of topics such as linking, epenthesis, deletion, reduction, and combinations of these processes, the pronunciation of words in discourse often is dramatically different from citation forms. Anne Wichmann looks at the functions played by English intonation in discourse, looking at the examples of *please*-requests, information structure, interaction management, and attitudinal meaning. Beatrice Szczepek Reed examines the behavior of prosody in discourse, especially the role of speech rhythm in managing interaction. Many aspects of communication are not tied to single phonological features but rather clusters of features. Finally, Ron Thomson looks at the meta-category of fluency and its relationship to pronunciation. Often thought to be directly related to some aspects of pronunciation, fluency is instead indirectly related to pronunciation but remains a topic that may be important for teaching.

The next section looks at the pronunciation of varieties of English. Initially, we hoped that the writers here would describe their varieties in terms of the international phonetic alphabet, believing that such a description would serve to highlight comparisons. Unfortunately, this proved to be much more difficult than we thought. Different traditions seem strongly entrenched in different areas of the English-speaking world, and each makes sense within its own native environment. Wells' (1982) use of key words, e.g., the *GOAT* vowel) often served as a unifying descriptive apparatus. As a result, each chapter has its own idiosyncrasies, but each is also very accessible. Each may require, however, greater familiarity with the IPA chart, especially to the different vowel symbols not often seen in descriptions of English. In addition, each general variety, such as Australian/New Zealand English, refers to a wide variety of regional and social dialects. Within the page limits, we asked authors not to focus on similarities within dialects, but rather to talk about socially significant pronunciations. The result is a catalogue of the richness of each variety.

Charles Boberg describes the pronunciation of North American English. A Canadian, Boberg is particularly well qualified to describe both Canadian and US pronunciations and to make sure that the dominance of US pronunciation does not overshadow the importance of Canadian English. Laurie Bauer (from New Zealand) provides the same kind of balance to the description of Australian/New Zealand English, demonstrating how the differences in the varieties were influenced by their earliest settlement patterns and differing immigration patterns. Clive Upton provides an abundant description of modern-day British English

pronunciation, including not only traditional RP but the geographic and social variety that defines English pronunciation in Great Britain and Ireland. Looking at South African English (the only variety seemingly without an -ing/-in' variation), Ian Bekker and Bertus Van Rooy describe fascinating L1 and L2 varieties of English and their connection to South Africa's social and historical development. As interesting and important as the native varieties of English are, nativized varieties of English have their own pronunciation patterns. Pramod Pandey's description of Indian English looks at perhaps the best described and most influential of these new Englishes. Like native varieties, Indian English has its own abundant regional and social variation. Finally, Cecil Nelson and Seong-Yoon Kang look at pronunciation through a World Englishes lens, giving a historical overview of a World Englishes view of English, and especially the role of pronunciation. In doing so, they demonstrate clear differences in approach between World Englishes approach and that of English as a Lingua Franca.

The next section is brief with only two chapters. It addresses the acquisitional issues for English pronunciation. Marilyn Vihman gives a state-of-the-art review of how English pronunciation is acquired by children as an L1. For those used to reading about L2 learning, this chapter will be eye-opening. For L2 pronunciation, Pavel Trofimovich, Sara Kennedy, and Jennifer Foote overview the important variables affecting L2 pronunciation development and provide questions for further research. The long-running debate about the differences between L1 and L2 acquisition has, by and large, not been strongly held for pronunciation learning. These two chapters should serve to show how distinct the two processes are.

The final section of the Handbook is the most directly relevant to teaching. In it, most papers address, explicitly or implicitly, questions of priorities and questions of students' cognitive engagement with pronunciation learning. Given limited time, which elements of pronunciation are most important and how should such decisions be made? Murray Munro and Tracey Derwing bring their considerable expertise to bear on how research insights into intelligibility can influence the teaching of pronunciation with an examination of current practice. Beth Zielinski looks at another issue in teaching, the long-running segmental/supra-segmental debate. The debate centers on the question of which is more important in the classroom, especially in situations where there is little time available for pronunciation teaching. Zielinski argues that the underlying assumption of the debate, that it is possible to separate segmentals and supra-segmentals, is faulty, and that both are essential. Graeme Couper brings a multidisciplinary approach to classroom research to bear on questions of teaching. He looks at what second language acquisition, social theories of learning, L2 speech research, and Cognitive Linguistics say in developing an approach to L2 pronunciation learning that is not defined primarily by what is currently done in the classroom.

In the next chapter, Robin Walker and Wafa Zoghbor describe an influential and sometime controversial approach to teaching English pronunciation, that of English as a Lingua Franca. This approach is based on Jenkins (2000) in which two NNSs of English are in communication with each other (an overwhelmingly common occurrence in the world today) and what kinds of pronunciation features are required for

them to be mutually intelligible. The approach was developed by Walker (2010) and is quite distinct from those pursued in most ESL and EFL contexts. In Intonation in Research and Practice: The Importance of Metacognition, Marnie Reed and Christian Michaud look at teaching intonation from a new perspective, that of metacognition. Intonation, even when it is taught, tends to focus on production, but the authors identify a difficulty with this approach. Students may successfully produce intonation in the classroom without understanding its communicative importance. As a result, they are unlikely to ever make what they have produced part of their own speech. Laura Sicola and Isabelle Darcy examine one of the most challenging yet recommended approaches to teaching pronunciation, the integration of pronunciation with other language skills. Wayne Dickerson, in the next chapter, argues for the importance of prediction in teaching pronunciation. Dickerson argues that predictive skills must be as important as perceptive and productive skills, and that predictive skills have a particular strength in empowering learners in pronunciation learning. Finally, Rebecca Hincks addresses technology, an area that is sure to grow and become even more influential in teaching pronunciation. She explains how speech technology works and explores how technology can be used to help learn pronunciation without and with automatic feedback, how it can evaluate pronunciation, and how it can provide automated speaking practice.

Single-volume handbooks are popular as reference sources. They offer a focused treatment on specialized topics that have a variety of interrelated topics that teachers and researchers are likely to understand inadequately. In an increasingly specialized profession, most teachers and researchers understand a few applied linguistics topics well, but there are many other topics with which they have only a passing acquaintance. English pronunciation is more likely than most topics to fit into the second category.

In summary, this *Handbook of English Pronunciation* is meant to provide:

- a historical understanding of the development of English pronunciation, the social role of accent, and the ways in which pronunciation has been taught over time;
- a description of some of the major varieties of English pronunciation and the social significance of pronunciation variants in those varieties;
- a description of the elements of English pronunciation, from sounds to syllables to word stress to rhythm to intonation;
- an examination of how discourse affects the pronunciation of segments and the meanings of supra-segmental features, as well as a discussion of pronunciation's connection to fluency;
- a discussion of how English pronunciation is acquired both in first and second language contexts and the variables affecting acquisition; and
- a selection of chapters that help to frame essential issues about how teaching pronunciation is connected to research and to the spread of technology.

One of the best things about editing this handbook has been learning that many of the things that we thought we knew were mistaken. Our authors come from

many countries and most of the continents, and many of them we had not had the pleasure of working with before starting this project. It is clear that brilliant work on English pronunciation is being done by extraordinarily talented and interesting researchers and teachers throughout the world. By bringing them together in one volume, we hope that you, the readers, will find many new and provocative ways to think about English pronunciation, and that you will find the handbook to be as interesting as we have in putting it together.

Marnie Reed and John M. Levis

REFERENCES

Abercrombie, D. 1949. Teaching pronunciation. *ELT Journal* (3)5: 113–122.

Breitkreutz, J., Derwing, T.M., and Rossiter, M.J. 2002. Pronunciation teaching practices in Canada. *TESL Canada Journal* 19: 51–61.

Burgess, J., and Spencer, S. 2000. Phonology and pronunciation in integrated language teaching and teacher education. *System* 28(2): 191–215.

Hinofotis, F. and Bailey, K. 1981. American undergraduates' reactions to the communication skills of foreign teaching assistants. *In: TESOL '80: Building Bridges: Research and Practice in Teaching English as a Second Language*, J. Fisher, M. Clarke, and J. Schachter (eds.), 120–133, Washington, DC: TESOL.

Jenkins, J. 2000. *The Phonology of English as an International Language*, Oxford: Oxford University Press.

Levis, J. (ed.). 2005. Reconceptualizing pronunciation in TESOL: intelligibility, identity, and World Englishes. *TESOL Quarterly* 39(3).

Macdonald, S. 2002. Pronunciation– views and practices of reluctant teachers. *Prospect* 17(3): 3–18.

Mugglestone, L. 2007. *Talking Proper: The Rise of Accent as a Social Symbol*, 2nd edition, Oxford: Clarendon Press.

Munro, M. and Derwing, T. 1995. Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning* 45: 73–97.

Smith, L. and Nelson, C. 1985. International intelligibility of English: directions and resources. *World Englishes* 4(3): 333–342.

Walker, R. 2010. *Teaching the Pronunciation of English as a Lingua Franca*, Oxford: Oxford University Press.

Wells, J.C. 1982. *Accents of English*, vol. 1, Cambridge: Cambridge University Press.

Part I The History of English Pronunciation

1 The Historical Evolution of English Pronunciation

JEREMY SMITH

Introduction

Since at least the nineteenth century, the study of sound-change has been at the heart of English historical linguistics and our current state of knowledge depends on the insights of generations of scholars. This chapter aims simply to give a broad outline of the current "state of the art", confronting basic questions of historical explanation. What does it mean to "account for" or "explain" a sound-change? How far can sound-changes be "explained"? How does one practise English historical phonology?

It is held here that historical phonology is as much history as phonology, and this insight means that evidential questions need to be addressed throughout. To that end, evidential questions are addressed from the outset. The chapter proceeds through the examination of a series of case studies from the history of English, ranging from the period when English emerged from the other Germanic dialects to become a distinct language to residualisms found in present-day varieties.

Overall, the chapter invites readers to reflect on their own practice as students of historical phonology; the explanations offered are, it is held here, plausible ones but by no means closed to argument. Good historiographical practice – for academic disciplines are of course collective endeavours – demands that such explanations should always be contested, and if readers can come up with better, more plausible explanations for the points made here, that is a wholly positive development, indicating new ways forward for the subject.

A question of evidence

Present-Day English is full of phonological variation; this variation, which is the outcome of complex and dynamic interactions across time and space, is valuable evidence for past states of English. To illustrate this point, we might take

The Handbook of English Pronunciation, First Edition. Edited by Marnie Reed and John M. Levis.

the varying British English pronunciations of the words (a) *good*, (b) *food*, and (c) *flood*: a Scot will commonly rhyme (a) and (b); speakers from northern England typically rhyme (a) and (c); southern British English speakers rhyme none of them. Another example: southern British English speakers have a phonemic distinction between /ŋ/ and /n/ in, for example, *sing*, *sin*; northern English speakers do not, since they retain a final plosive in *sing* and for them [ŋ] is environmentally conditioned (and thus an allophone of, and not a distinct phoneme from, /n/). Many speakers of Scots, the traditional dialect and accent of Scotland, as well as speakers from north-east England, will pronounce the vowels in words such as *cow*, *now*, *house* with a close rounded back monophthong rather than (as southern speakers do) with a diphthong (see further Wells 1982).

Those learning to read, or non-native speakers, might reasonably expect, in a supposedly *phonographic* language such as English, that words ending in the same three letters, viz. *–ood*, in the written mode, should rhyme when read aloud, but, as we have just observed, in many accents of English they do not. The reason for the variation, and for the mismatch between spelling and sound, is that sound-changes have occurred since the spelling-system of English was established and standardized, and that these sound-changes have *diffused* differently through the lexicon in different parts of the English-speaking continuum. Some changes have only been adopted in some varieties.[1]

The outcome of such patterns of divergence and diffusion is a body of *residualisms*, i.e., older forms of the language that remain in some accents but have ceased to be used in others (see Ogura 1987, 1990; Wang 1969; Wells 1982). The Scots/north-eastern English monophthongal pronunciations, for instance, of *cow*, *now*, *house* reflect the monophthongal pronunciation that seems to have existed in English a thousand years ago, cf. Old English *cū*, *nū*, *hūs* respectively. These pronunciations are therefore residualisms.

Residualisms are one of the major sources of evidence for the *reconstruction* of past states of pronunciation. We might illustrate the process of reconstruction using residualisms by comparing the British, Australian, and US pronunciations of the word *atom*; British and Australian speakers pronounce the medial consonant as /t/ whereas US speakers characteristically use a voiced alveolar tap, meaning that in US English the word *atom* is a homophone with *Adam*. It is usual to consider the US pronunciation to be an innovation, whereas the other usages are residualisms, the evidence for this interpretation being that US speakers characteristically voice intervocalic sounds in derived forms, cf. US English intervocalic /d/ (however precisely realized) in *hitter* beside final /t/ in *hit*, beside /t/ in both environments in British and Australian usage. Such reconstructive processes are, of course, the basis of comparative linguistics.

However, deciding what is a residualism and what is not can be a difficult matter without further information. To take a large-scale example: the phenomenon known as Grimm's law (the "First Consonant Shift"), whereby a series of consonants in the Germanic languages seem to have undergone a comprehensive redistribution within the lexicon, is traditionally described as a Germanic innovation. Illustrative examples are given in Table 1.1.

Table 1.1 Grimm's law cognates in Germanic and non-Germanic languages.

	Germanic examples	*Non-Germanic examples*
/f/ - /p/	English *fish*, Norwegian *fisk*	Latin *piscis*, French *poisson*, Welsh *pysg*
/θ/ - /t/	English *three*, Icelandic *þrír*	Latin *trēs*, French *trois*
/h/ - /k/	English *hound*, German *Hund*	Latin *canis*, Welsh *ci*, Tocharian *ku*

However, some scholars, arguing that a similar process is also found in Armenian, like Germanic a "peripheral" language within the Indo-European group but at the eastern as opposed to the western end of that language-family's extent, have argued that Grimm's law represents a residualism rather than an innovation. This so-called "*glottalic" theory* is highly controversial, but that it has found purchase with at least some scholars indicates the nature of the problem (see Smith 2007: ch. 4).

The study of residualisms as evidence for the history of pronunciation, therefore, is – where possible – combined by researchers with other sources of evidence: sound-recordings, available since the end of the nineteenth century; contemporary comments on past pronunciation; past spelling-practices, given the mapping between speech and writing found in phonographic languages; and the practices of poets, in terms of rhyme, alliteration, and metre. Taken together, these various pieces of evidence allow scholars to develop plausible – though never, of course, absolutely proven – accounts of past accents, and sometimes even to offer plausible explanations for how particular accentual features emerged. A series of case studies follows, with special reference to the history of English, to illustrate the process of developing such plausible accounts and explanations.

Case study 1

Voiced and voiceless fricatives: development of new phonemic categories

The first of these case studies deals with the Present-Day English phonemic distinction between voiced and voiceless fricatives, a distinction that has emerged during the history of English and is reflected – albeit sporadically and unevenly – in Present-Day English spelling. The example also allows us to ask a certain key, and surprisingly neglected, question: what is a sound-change?

One such distinction, which often puzzles present-day learners of English, is to do with the pronunciation of the word *house*; when used as a verb, the word ends with /z/ but, when used as a noun, it ends with /s/. The usual historical explanation is as follows: in Old English, voiceless [s] and voiced [z] were allophones of the same phoneme, conventionally represented by /s/, and therefore in *complementary distribution* within the sound-system. It seems that /s/ was pronounced

voiced intervocalically, but voiceless when a word-final. The Old English word for "house" (noun) was *hūs*, while the Old English word for "house" (verb) was *hūsian*; when, in the transition from Old to Early Modern English, inflectional endings such as *–ian* were reduced and ultimately lost, a voiced sound emerged in final position in words such as "house" (verb), leading to the current pattern for the sound's deployment. Since "house" (noun) and "house" (verb) now have distinct meanings marked by replacement of single word-final segments, the two words have come to form a *minimal pair* for the purposes of phonological analysis, and the phonemes /s, z/, now in *contrastive distribution*, may thus be distinguished.

Of course, the evidence we have for the initial complementary distribution can only be deduced; direct evidence, in the form of contemporary commentary or distinctive spellings from Old English times, is almost entirely lacking and the distribution of forms means that poetic evidence is not to be had. The issue is one of plausibility, in that the process of *phonemicization* just described aligns with known developments elsewhere in the linguistic system, notably inflectional loss.

Spelling evidence for sound change is really only available on a large scale from the Middle English period. Middle English is notoriously the period in the history of English when there is a closer alignment between spelling and pronunciation than before or since. Written English had a parochial rather than national function, used for initial or otherwise restricted literacy, while – following Continental practice – unchanging, invariant Latin was deployed as the language of record across time and space. Thus it made some sense to reflect English phonological variation in the written mode, since that made teaching reading easier. Only when English, towards the end of the medieval period, took on the role of a language of record did variation become inconvenient. The *standardization* of written English was a formal response to a change in linguistic function. That English spelling could remain fixed while pronunciation changed was first discussed by Charles Butler in his *English Grammar* (1633), who saw the development as regrettable and thus needing reform (Dobson 1968: 165), but the socially useful functionality, for record-keeping purposes, of a fixed spelling-system, despite a phonographic mismatch between spelling and widely attested pronunciations, has meant that comprehensive spelling-reform in English has never succeeded.

It is therefore possible – at least sometimes – to see reflections of sound-change in changes in spelling. As with the [s]/[z] distinction, Old English made no phonological distinction, it seems, between voiced and voiceless labio-dental fricatives and as a result the spelling <f> was used to reflect both, e.g., *fela* "many", *hlāf* "loaf" (both with [f]), but *yfel* "evil" (with medial [v]). A phonological distinction seems to have emerged in the Middle English period largely as a result of the adoption of loan-words from French, e.g., *fine*, *vine*, and this distinction became sufficiently salient for a spelling-distinction, between <f> and <v>, to be adopted and even extended to native words, such as *evil*. The <f>/<v> distinction first emerged in Middle English and has been sustained ever since.

However, it is noticeable that even in Middle English conditions such developments do not always follow. Distinctions between other voiced and voiceless

fricatives, i.e., the alveolars /s, z/ (as we have just seen) and the dentals /θ, ð/, also emerged, but the spelling-evidence for such developments is uncertain. The letter <z> remains marginal in Present-Day English spelling, used in the initial position only in exotic words such as *zoo*, *zebra* and even replaced by other letters altogether in *xylophone*, *xerox*; in medial and final positions it is also in some sense "optional", cf. the variation between *criticise*, *criticize*, or the fact that the word *ooze* is a homophone with the river-name *Ouse*. For Shakespeare, <z> was an "unnecessary letter" (*King Lear* II.2) and in Middle English <z> is witnessed only sporadically. It is noticeable that the only texts to use <z> consistently in the initial position are Middle Kentish ones, such as the *Ayenbite of Inwyt*, surviving in a manuscript localized to Canterbury in 1340, where a consistent distinction is made between, for example, *zom* (from Old English *sum* "a certain") and *som* (from Old French *sum* "a sum (of money, etc)". Initial voicing of fricatives seems to have survived in Kentish until the end of the nineteenth century though is now recessive (see Smith 2000 and references there cited).

Similarly marginal is the distinction in voiced and voiceless dentals. Present-Day English deploys <th> for both /θ/ and /ð/, except in specialist vocabulary such as *sandhi* or in forms made up for literary effect by philologists, such as the name *Caradhras* in J.R.R.Tolkien's *The Lord of the Rings*; in both cases <dh> represents the voiced fricative sound. The reason for this limited reflection of a phonological distinction seems to be that there is only a limited set of minimal pairs, e.g., *thy*, *thigh*, and *that*, and at least in the initial position, the voiced dental fricative is restricted to "grammar words" such as *the*, *that*, *this*, *those*, *these*, *there*, *though*, or in certain pronouns such as *they*, *them*, *their*. In Middle and Early Modern English texts, there is some evidence that some scribes deployed <þ> – sometimes written in a manner indistinguishable from <y> – only in such words (e.g., the common use of <ye> for "the"). Such practice may reflect a sound-distinction, but equally plausibly it could be argued that it is simply a space-saving device, whereby a form largely predictable from context could be represented in abbreviated fashion (the custom of abbreviating forms such as "the" or "that" as <ye> or <yt>, with superscript second letters, would support the latter interpretation).

The key point, of course, is that there is no necessary connection between what a medieval or renaissance scholar would have called the *figura* (written manifestation of a *littera* "letter") with a particular *potestas* (sound-equivalent) (see Abercrombie 1949). To demonstrate this point, we might take, for instance, spellings of the words "shall", "should", common in the Middle English of Norfolk, viz. *xal*, *xuld*. In such cases, it is notoriously hard to establish the potestas of <x>. Is <x> in such words simply a local spelling for [ʃ] or does it represent a distinct sound? Its restriction to the words "shall", "should" (until the very end of the Middle English period, when it is sporadically transferred to words such as *xuldres* "shoulders") would suggest the latter, but there is no certainty as to the precise potestas to be assigned to it.

Support for a voiced/voiceless distinction in the fricatives, at least for the alveolar and dental sets, is suggested rather than proven by the spelling-evidence, and

other information is needed if we wish to establish the phonemicization in the history of English pronunciation. Unfortunately, there is no meaningful discussion of English pronunciation until the sixteenth century, when English became a respectable subject for intellectual study rather than simply a "vulgar" tongue; however, the evidence from then on becomes full. John Wallis's *Grammar of the English Language* (1653), for instance, noted the distinction between what he called "hard *s*" and "soft *s*", in which the latter was pronounced "per *z*" in *a house*, *to house* respectively (Kemp 1972: 178–179), and Wallis regretted the failure in English spelling to distinguish voiced and voiceless dental fricatives, which he regarded as "an unfortunate practice" (Kemp 1972: 176–177). Wallis states that the Welsh use <dd> for the voiced sound "though some maintain that *dh* would be a better way of writing it than *dd*; however they have not succeeded in getting the old established custom altered" (Kemp 1972: 177).

Interestingly, the labio-dental voiced/voiceless distinctions are not discussed to the same extent, possibly because the spelling-distinction was already accepted by early modern times. The spelling *hlīuade* for the third-person preterite singular of *hlīfian* "stand tall, tower" appears in the late tenth century *Beowulf Manuscript* (MS London, British Library, Cotton Vitellius A.xv, *Beowulf* line 1799), beside the more common *hlīfade*. The spelling with <u> is usually taken as the earliest instance of an attempt to reflect a voiced–voiceless distinction in English spelling.

A good working definition of sound-change might be as follows:

> *Sound-change is a phenomenon whereby speakers adjust their phonologies, or sound-systems. The raw material for sound-change always exists, in the continually created variation of natural speech, but sound-change only happens when a particular variable is selected in place of another as part of systemic regulation. Such processes of selection take place when distinct systems interact with each other through linguistic contact, typically through social upheavals such as invasion, urbanization, revolution, or immigration.*

However, two issues become fairly clear from the discussion so far. Firstly, as the form *hlīuade* and the current restricted distribution of the voiced and voiceless dental fricatives suggest, sound-change is what might be termed an *emergent* phenomenon. That is, sound-changes are not sudden affairs but typically diffuse through time and space in a *"sigmoid-curve" pattern*, working their way through the lexicon. Diachronic discussion is not a matter of aligning a series of synchronic descriptions of phonological inventories at given points in time, i.e., a series of "maps". It is a different kind of discourse (for the notion and importance of emergence, see especially the essays in Bybee and Hopper 2001).

Secondly, it is clear that, although almost all scholars accept a general narrative about the history of voiced and voiceless fricatives in the history of English, the evidence is indicative rather than conclusive. Potestates map on to figurae, but in complex ways, and without access to recorded sound from any period before the end of the nineteenth century it is not possible to offer any final, demonstrable proof of the structure of past sound-systems. The argument, as so often in historical study, is based on the *plausible interpretation* of fragmentary indicators.

Digraphs and diphthongs

The previous section focused on what is arguably the major phonological development in the history of English sounds: the emergence of a whole distinct category of phonemes. Changes in English vowels are more widespread, but making evidence a starting-point can also be most illuminating.

As with consonantal change, that potestates map on to figurae in complex ways can be illustrated with reference to the history of English vowels, and a Present-Day English example makes the point. In most modern accents, words with <ee> and <ea> commonly rhyme, e.g., *meet*, *meat*, although there are of course numerous exceptions, e.g., *greet*, *great*, and some alternative rhyming patterns, commonly, where the vowel is followed by /r/, e.g., *pear*, *pair* rather than *pear*, *peer* (although cf. the non-rhyming *fear*, *fair*), or by a dental or alveolar consonant, e.g., *breath* (rhyming with the personal name *Seth*) and *dead* (rhyming with *bed*). In some varieties, particularly conservative ones, what are clearly older patterns survive residually, e.g., in some accents of Irish English *meat* rhymes with *mate* rather than *meet*. The current complex distribution of <ea> spellings in relation to sound-systems is the result, as we might expect from the discussion so far, of sound-changes diffusing incompletely and irregularly across the lexicon subsequent to the standardization of the writing system.

It might be expected, in periods before the writing system became standardized, that the relationship between figurae and potestates might be closer, i.e., the language-variety in question would be more completely *phonographic*. However, despite a tradition of research of more than a century, very basic problems in the interpretation of vowel-potestates remain contested by scholars.

Anglo-Saxonists, for instance, still debate the existence of basic phenomena such as the nature of the diphthongal system and the interpretation of the spellings <ea, eo, ie>. Questions asked, still not conclusively answered, include:

1. Do these spellings really represent diphthongs?
2. Are they to be seen as equivalent to long monophthongs, i.e., VV?
3. How far are (as conventional wisdom holds) the "short diphthongs" <ea, eo, ie> to be seen as metrically equivalent to short vowels, i.e., V (vowels with which, historically, they tend to merge)?
4. How are the individual elements within these diphthongs (if that is what they are) to be pronounced?

These questions form a major conundrum in the study of Old English phonology.

Almost all scholars accept the existence in the West Saxon dialect of Old English of the long diphthongs spelt <ea, eo>, which represent the reflexes of Germanic diphthongs as well as the products of certain sound-changes. These diphthongs were "bimoric", i.e., VV in terms of *metrical weight*, and thus equivalent to long monophthongs, sounds with which historically they tended to merge. The problem arises with the so-called "short diphthongs", which were not the reflexes of Germanic diphthongs but arose as the result of sound-changes such as breaking or

"palatal diphthongization", and have been believed by many scholars to be monomoric, i.e., V, and thus equivalent in metrical weight to a short monophthong. Richard Hogg sums up this view as follows: "… the traditional position holds that <ea, eo, io> always represented diphthongs both long and short except where the orthographic evidence suggests otherwise or the linguistic development is implausible …" (1992: 17). The key problem is, as David White has pointed out (2004: *passim*), that such short diphthongs are vanishingly rare in world languages, and indeed not found in living languages at all; their presence in standard descriptions is the outcome in all cases of scholarly reconstruction.[2]

One argument offered originally by Marjorie Daunt (1939, 1952) and reiterated by White (2004) is that spellings such as <ea, eo>, when representing the "short diphthongs", include a diacritic element, flagging the quality of the following consonant. Certainly it is generally accepted that such diacritic usages occur in Old English, e.g., spellings such *sēcean* "seek" (beside more common *sēcan*), or *geong* "young" (which would have yielded Present-Day English ******yeng* if <eo> in this word had represented one of the presumed "short diphthongs"). It could therefore be argued that <ea, eo> in words such as *eald* "old", *earn* "eagle", *weorpan* "throw", *eolh* "elk" represent /æ/ or /e/ followed by a "back (i.e., velarized) consonant"; <eo> in *heofon* "heaven" would be an attempt to represent /e/ "colored" by the back vowel in the unstressed syllable. Daunt pointed out that digraphs of various kinds were deployed by Old Irish scribes to flag the quality of neighboring consonants, and Old Irish scribal practice strongly influenced Old English usage.

However, there are problems with this analysis. Minimal pairs arose in West Saxon, subsequent to the operation of the sound-change that produced <ea> in *eald*, *earn*, etc., which seem to indicate that <ea> was perceived in West Saxon as distinct in quality from <æ>, e.g., *ærn* "house" beside *earn* "eagle"; despite suggestions to the contrary (e.g., White 2004: 80), it seems likely that, in the conditions of vernacular literacy obtaining in West Saxon, this difference indicates a real distinction in pronunciation. If there were no difference in pronunciation we would expect variation in spelling between ******æld* and *eald* in West Saxon, and such a variation does not occur.

Although some languages (e.g., Scottish Gaelic) have a three-way length distinction, viz. V, VV, VVV (see Laver 1994: 442), it seems unlikely that Old English had the same system, with the short diphthongs to be interpreted as bimoric (VV) and the long diphthongs as trimoric (VVV). The "long diphthongs" of OE derive in historical terms from bimoric (VV) Proto-West Germanic diphthongs, and there does not seem to be any good reason to posit a lengthening, especially as, in later stages of the language, they tend to merge with long monophthongs (VV).

Perhaps the most economical explanation would be to see the "short diphthongs" as consisting of a short vowel followed by a so-called glide vowel, i.e., Vv in the environment of a following back consonant. Daunt herself argued that "there was probably a glide between the front vowel and the following consonant" (Hogg 1992: 18–19, and see references there cited). The distinction between monophthongs plus glides and diphthongs is a tricky one, but recent

experimental work on Spanish suggests that a robust distinction is possible (see Hualde and Prieto 2002). The spelling <ie> is used in Early West Saxon to represent the outcome of further sound-changes that affected <ea, eo>, and it therefore seems logical – if the Daunt/White interpretation is accepted – to assume that it, too, represents a diphthong, probably of the same kind (i.e., full vowel plus hiatus vowel).

Establishing the sound-equivalent (potestas) of a particular spelling (figura) is one thing: proceeding to explain the conditions under which a particular potestas emerged is another, and here we are on even more tenuous ground at such an early date in the history of English. The Old English spelling <ea> in *eald*, *earn*, etc., is a product of the sound-change known as "*Breaking*", usually defined as a diphthongization in the environment of a following "back" (i.e., velar) consonant. Whether <ea> is to be interpreted as a diphthong or not is, as we have just seen, a complex question, but all scholars agree that the consonants <l, r>, etc., are "back" in terms of the Old English system. The question is, though, when did they *become* back consonants to induce the change?

One plausible possibility is that the precise realization of <l> in the Old English dialects manifesting breaking had undergone a change as the result of contact with other varieties, a change in consonantal realization that had a knock-on effect on the pronunciation of the preceding vowel. It is thus relevant to refer back to consonantal change when accounting for the evolution of vowels, flagging the dynamic interconnectedness of sound-changes. Breaking is the first sound-change that can be clearly located in Anglo-Saxon England after the so-called *Adventus Saxonum* ("the coming of the Saxons"), the period of transition between Romano-Celtic Britain and Anglo-Saxon England; earlier sound-changes, e.g. "*First Fronting*" (sometimes known as "Anglo-Frisian Brightening"), date from the period when the Angles and Saxons were still on the Continent of Europe. It thus developed, in West Saxon, at a time when Saxons were coming into contact with Angles in a condition of confused and complex social ties.

There is some evidence that, in Old Anglian, /l/ and /r/ were back consonants. Old Anglian was in origin the variety furthest north within the West Germanic-speaking area, being spoken in the area immediately abutting the most southern varieties of North Germanic, and the continual interchange between North and West Germanic, often commented on by linguists (see for instance Haugen 1976: *passim*), would clearly have impacted most upon it. Many of these southern varieties even now have a "dark /l/", often referred to as "thick" or "cacuminal" /l/. It could therefore be argued that, when Anglian and Saxon varieties came into contact with each other as a result of the Adventus Saxonum, Saxons attempted to reproduce Anglian usage in situations of language contact; a "dark" form of /l/ would result. That Saxons would have imitated Anglians rather than vice versa is suggested by the evidence – admittedly somewhat tenuous – that Anglians dominated the early Anglo-Saxon polity: after all, the name "England" derives from "Angle", and the name "Saxony" is applied to an area of present-day Germany (see further Smith 2007: ch 4, and references there cited).

The Great Vowel Shift

In the previous section, the explanation offered for change was in some sense sociolinguistic, but there were limits to such an approach, derived, quite simply, from the comparative paucity of evidence. The best that can be hoped for from such explanations is plausibility linked to certain arguments to do with similarities between past and present. In this section, greater evidence allows us to make such arguments more convincingly.

Such explanations as that just offered for the origins of Breaking, as the result of language contact in situations where one group might be considered more prestigious than another, may be tenuous, but they gain traction from the observable fact that such situations are observable in present-day language. As William Labov famously argued in what may be considered a foundational statement of the subdiscipline of *historical sociolinguistics*, the present can be used to explain the past (Labov 1974). Since the so-called "*uniformitarian hypothesis*", accepted by linguists, holds that speakers in the past – like us – reflected their social structure in language (see, for example, Romaine 1982 and Machan 2003), it seems unarguable that the social setting of language-use in early times had an effect on linguistic development, specifically sound-change. The tenuousness of the explanation relates to the difficulty not of the principle but of our limited understanding of the precise social circumstances that obtained at the time.

It is therefore arguable that the more information we have about social structure the higher degree of plausibility there is about explaining a given sound-change. Thus a later change, such as the *Great Vowel Shift* of the fifteenth and sixteenth centuries, a process of raisings and diphthongizations that distinguishes the phonologies of Late Middle English period from those of the Early Modern English period and that may be described as a redistribution of sounds within the lexicon, can be explained fairly convincingly as the outcome of interaction between social groups in conditions of increasing urbanization.[3]

The origins of the Great Vowel Shift have, notoriously, been regarded by many scholars as "mysterious" (Pinker 1994: 250), an adjective that would seem to close down discussion. However, an interest in the Shift's origins has persisted, particularly amongst scholars whose work engages with sociolinguistic concerns.

It is noticeable that the Shift took place at a key moment of transition in the history of English, when English ceased to be a language of comparatively low status in comparison with Latin and French and began to take on national roles, i.e., it underwent a process that Einar Haugen has referred to as *elaboration* (Haugen 1966; cf. also Hudson 1980: 32–34, and references there cited). The elaboration of English meant that prestigious varieties of that language began to emerge. The story of the Southern Great Vowel Shift relates, I have argued, intimately to that emergence. It seems that the Southern Shift derives from sociolinguistically-driven interaction in late medieval/early Tudor London, whereby socially mobile immigrant groups *hyperadapted* their accents in the direction of usages that they perceived as more prestigious. Such a process can be paralleled

in modern situations, whereby linguistic innovation is located in the usage of those who are weakly tied to their social surroundings (see Milroy 1992).

The origins of the Southern Shift correspond in date to four major – and, I would argue, linked – developments in the external and internal history of the English language. These developments are as follows:

a. *The rise of a standardized form of English.* At the end of the fourteenth and the beginning of the fifteenth centuries, it is possible to detect, in the written mode and to a lesser extent in speech, the emergence of focused forms of language that are the precursors of Present-Day "standard" varieties.
b. *The growth of London.* The end of the Middle Ages and the beginnings of the Tudor period saw the increasing significance of London as England's major administrative and trading centre. From the fourteenth century onwards there was a major influx of immigration into the capital from the countryside as folk sought to improve their condition in the city. This is the age of the quasi-mythical figure of Dick Whittington, who moved to London, where the streets were (it was said) paved with gold, to make his fortune. The result was that London became, according to contemporaries, the only English city comparable in size and importance to continental centers such as Paris, Venice, and Rome (see, for a convenient account, Ackroyd 2002, and references there cited). London society, which (as nowadays) attracted incomers from elsewhere eager to take advantage of the opportunities it had to offer, may be characterized as one with *weak social ties* in comparison with those which obtained in the much more stable, less dynamic village society that existed elsewhere in England.
c. *The loss of final –e*. The Shift corresponds in date to a grammatical development of considerable prosodic significance: the development of what is essentially the Present-Day English grammatical system with *the loss of inflectional –e*. Final *–e* was still in use in adjectival inflections in Chaucer's time, as established (*inter alia*) by the poet's verse practices, but the generations that followed Chaucer, from the end of the fourteenth century onwards, no longer recognized the form. The loss of *–e* had major implications for the pronunciation of English, whose core vocabulary became, to a large extent, monosyllabic in comparison with other major European languages.
d. *Phonemicization of vowels affected by Middle English Open Syllable Lengthening in those accents where these vowels did not undergo merger.* This development was a consequence of the loss of final *–e*. There is good evidence, from contemporary rhyming practice in verse, that the comparatively prestigious form of speech represented by that of Geoffrey Chaucer distinguished carefully between the reflex of Old English *e* and *o*, which had undergone a quantitative change known as *Middle English Open Syllable Lengthening* and the reflex of Old English *ēa*, *ǣ*; with the loss of final *–e*, this distinction became phonemicized in Chaucer's (more properly, Chaucer's descendants') variety and thus perceptually salient. However, in other varieties outside London, Middle English Open Syllable Lengthening-affected *e*, *o* merged with the reflexes of Old English *ēa*, *ǣ*, and *ā* >$\bar{Q}$ respectively. These two systems may be characterized as System I and System II respectively.

With the rise of London and the perception of there being a prestigious form of speech that coincided with it, users of System II, whose social situation may be characterized as weakly tied, came into contact with users of System I. System I speakers distinguished phonemically between Middle English Open Syllable Lengthening-affected *e* and *o* and the reflexes of Old English *ēa*, *ǣ*, and *ā* > *Q̄*, whereas System II speakers did not. Moreover, it seems likely that System I speakers, with a habit of pronouncing much of their stylistically marked vocabulary in a "French" way – see (a) – would have distinct ways of pronouncing mid-close *ē* and *ō*; there is some evidence that French *ē* and *ō* were realized as somewhat higher in phonological space than the reflexes of English *ē* and *ō*, and adoption of French-influenced usages would have been encouraged by the presence of the extra phoneme, derived from Middle English Open Syllable Lengthening, in both front and back series of long vowels. R.B. Le Page has suggested that the aristocracy of the late fourteenth and fifteenth centuries were likely "to adopt affected forms of speech as a means of 'role-distancing' from the lower classes, from whom they had hitherto been differentiated by speaking French" (cited in Samuels 1972: 145–146). Further, if the raised "French" style pronunciations of *ē* and *ō* were adopted by System I speakers, it seems likely that diphthongal pronunciations of the close vowels *ī* and *ū*, which are attested variants within the phonological space of close vowels in accents with phonemic length, would have been favored by them, viz. [ɪi, ʊu], in order to preserve distinctiveness. Such a development would mean that a four-height system of monophthongal long vowels would be sustained, with Middle English /i:/ being reflected as a diphthong, albeit one with a comparatively close first element.[4]

We would expect in such circumstances that hyperadaptations would follow, and this is the basis of the argument for the origins of the Shift offered here. System II speakers, who may be characterized as weakly tied, socially aspirant incomers, encountered System I speakers whose social situation they wished to emulate. The process, it might be plausibly argued, would have worked somewhat as follows. System II speakers would have heard System I speakers using what they would have perceived as a mid-close vowel in words where they would use a mid-open vowel. Since final *–e* had been lost there would not be a grammatical rule to identify when such vowels should be used, and System II speakers, who formed the rising class of late medieval and early Tudor London, would replace their mid-open vowels (whether derived from Middle English Open Syllable Lengthening-affected *e*, *o* or from Old English *ēa*, *ǣ*, and *ā* > *Q̄*) with mid-close ones. There would be phonological space for them to do so since they were also attempting to imitate the socially salient raised allophones of System I speakers' "French" style raised /e:, o:/. Since these latter pronunciations were themselves not in the inventory of System II speakers, it seems likely that such pronunciations were perceived as members of the phonemes /i:, u:/ and would be reproduced as such (on hyperadaptation, see Smith 2007, and references there cited, especially Ohala 1993).

Of the remaining developments in the Shift, diphthongization of front vowels would derive from attempts by System II speakers to imitate System I speakers' [ɪi, ʊu] allophones of /i:, u:/. Such selections would be encouraged by the need to

retain perceptual distance from the "French" style raised /e:, o:/, hyperadapted by System II speakers as /i:, u:/. As I have suggested elsewhere, the later development whereby Middle English /a:/ > /ɛ:/ probably derives from a distinct, sociolinguistically-driven process. Middle English phonemic /a:/ was comparatively new in most Southern English accents, being derived largely from Middle English Open Syllable Lengthening-affected /a/. The main accent in the South-East where phonemic /a:/ had existed beforehand was the Essex dialect, which seems to have been the "old London" usage characteristic of low-prestige speakers in the area. A raised pronunciation of Middle English /a:/, probably as [æ:], would have been another way of marking social distinction, which System I speakers would have been keen to make. System II speakers, attempting to replace their own realizations of /a:/ with System I's [æ:], would have tended again to overshoot, identifying the System I [æ:] pronunciation with the next phoneme in their own series, viz. /e:/.

The outcome of all the developments just described was the distribution of vowels attested by the best writers on pronunciation in the sixteenth century. The developments just argued for, incidentally, also illustrate how sound-change is a processual, emergent phenomenon, not something that suddenly appears in saltatory fashion, as might sometimes appear to be the case from handbook accounts.

Explaining sound-change

We might now move to central issues raised by the case studies discussed. Historical explanations, such as those just provided for Breaking and the Great Vowel Shift, are necessarily exercises in plausible argumentation, and a plausible argument is not absolutely proven. In historical subjects, absolute proof is not to be had. The question, therefore, is: how can we assess the success of an historical explanation?

As I have argued elsewhere (Smith 2007), certain historical approaches, e.g., postmodernism, have emphasized the "observer"s paradox", the way in which the frame of reference of the investigator constrains the enquiry. However, as I have suggested, the observer's paradox should not be seen as disabling, but rather it places certain ethical requirements on historians: to be self-critical, to be open to other interpretations of events, and (above all) to be humble. Historians are (or should be) aware that their work is in no sense a last word on a topic but simply part of a continuing discussion in which their views may eventually come to be displaced. Explanations of sound change, like all historical explanations, are successful if they meet certain criteria of plausibility. As April McMahon has put it, "we may have to accept a … definition of explanation at a … commonsense level: explanation might … constitute 'relief from puzzlement about some phenomenon'" (1994: 45, and references there cited).

In assessing the plausibility of the accounts of the Shift just offered, it is perhaps a good idea to return to the notion of the uniformitarian principle, a notion that underpins what is probably the most fruitful current development in the study of

the subject, viz. historical sociolinguistics (see further Millar 2012 and references there cited), and a renewed focus on what has been called the "linguistics of speech". Such a *parole*- (as opposed to *langue*)-based approach to linguistic investigation is informed by the close analysis of large bodies of data, both from the present-day and from the past, harnessing insights about the "dynamic" nature language derived from complexity science (for which see most importantly Kretzschmar 2009). The linking of present-day and past circumstances – as flagged by Labov back in 1974 – is crucial; if sound-changes in present-day circumstances take place because of certain social conditions, and if the phonetic processes that obtain in those circumstances (i.e., hyperadaptation) may be observed, then it seems at least plausible that similar processes governed sound-changes in the past. The study of past sound-changes, therefore, is a project that must be linked closely to an understanding of the dynamic and complex processes of social history. In so doing, we may be "relieved from puzzlement" – which is, in English historical linguistics, probably as good as it gets.[5]

NOTES

1 In a phonographic language there is, broadly speaking, a mapping between grapheme and phoneme. A *logographic* language, by contrast, is one where the mapping is between grapheme and notion. Written versions of Western European languages are largely phonographic; written Chinese is logographic. The difference may be illustrated by the symbols used for numbers; "8" is a logograph, corresponding to the written/spoken usages *eight* (in English), *huit* (in French), *otto* (in Italian), *acht* (in German), or indeed the spoken usages *bā* (Mandarin Chinese), *takwas* (Hausa), *siddeed* (Somali), or *walu* (Fijian). There are advantages to logographic languages; German speakers may not be able to understand Fijian speakers when they write in their native languages, but both Germans and Fijians will be able to understand each other's mathematical symbols. Famously, Cantonese and Mandarin are not mutually intelligible when spoken, but since the writing-system commonly deployed in varieties of Chinese is in principle logographic it is possible for users of these varieties to understand each others' writings. Logographic systems are problematized by their use of a very large number of symbols, and they are thus a challenge to the memorizing powers of those learning to read and write, but it is undeniable that they are useful as a language of record and transaction – which is why they emerged in Imperial China.

2 Richard Hogg was of course aware of the difficulty, although – appropriately in a handbook – he tended to the conventional view, and his qualification is therefore carefully expressed. A fuller quotation reads: "… the traditional position holds that <ea, eo, io> always represented diphthongs both long and short *except where the orthographic evidence suggests otherwise or the linguistic development is implausible* …" (1992: 17; my italics).

3 Five-height systems of monophthongal phonemes are attested in the world's languages, but are rare; three- and four-height systems are much more common (see Maddieson 1984: *passim*).

4 As well as a "full" Shift affecting both the long front and long back vowels of Middle English, characteristic of southern varieties, there was also a distinct Shift, affecting

primarily long front vowels, which is found in Northern accents. The discussion in this chapter focuses on the "full" or Southern Shift; for a discussion of both in much more detail, see Smith 2007: ch. 6, and references there cited. It is argued that the triggering of the "Northern" Shift was the result, like the Southern Shift, of socially-driven linguistic choices (i.e., it was a sociolinguistic phenomenon), whose outset related to earlier shifts in the back series of long vowels consequent on interaction with Norse.

5 For a similar attempt to use the present to explain the past, but with reference to a much more archaic set of sound-changes, see Jane Stuart-Smith's discussion of the processes involved in ancient Italic accents (2004).

REFERENCES

Abercrombie, D. 1949. What is a "letter"? *Lingua* 2: 54–63.

Ackroyd, P. 2002. *London: The Biography*, London: Faber.

Bybee, J. and Hopper, P. (eds.) 2001. *Frequency and the Emergence of Linguistic Structure*, Amsterdam: Benjamins.

Daunt, M. 1939. Old English sound changes reconsidered in relation to scribal tradition and practice. *Transactions of the Philological Society* 108–137.

Daunt, M. 1952. Some notes on Old English Phonology. *Transactions of the Philological Society* 48–54.

Dobson, E.J. 1968. *English Pronunciation 1500–1700*, Oxford: Clarendon Press.

Haugen, E. 1976. *The Scandinavian Languages*, London: Faber and Faber.

Haugen, E. 1966. Dialect, Language, Nation. *American Anthropologist* 68: 922–935.

Hogg, R. 1992. *A Grammar of Old English I: Phonology*, Oxford: Blackwell.

Hualde, J.I. and Prieto, M. 2002. On the diphthong/hiatus contrast in Spanish: some experimental results. *Linguistics* 40: 217–234.

Hudson, R. 1980. *Sociolinguistics*, Cambridge: Cambridge University Press.

Kemp, J.A. (ed.) 1972. *John Wallis's Grammar of the English Language*, London: Longman.

Kretzschmar, W. 2009. *The Linguistics of Speech*, Cambridge: Cambridge University Press.

Labov, W. 1974. On the use of the present to explain the past. In: *Proceedings of the 11th International Congress of Linguists*, L. Heilmann (ed.), 825–851, Bologna: Il Mulino.

Laver, J. 1994. *Principles of Phonetics*, Cambridge: Cambridge University Press.

McMahon, A. 1994. *Understanding Language Change*, Cambridge: Cambridge University Press.

Machan, T.W. 2003. *English in the Middle Ages*, Oxford: Oxford University Press.

Maddieson, I. 1984. *Patterns of Sounds*, Cambridge: Cambridge University Press.

Millar, R.M. 2012. *English Historical Sociolinguistics*, Edinburgh: Edinburgh University Press.

Milroy, J. 1992. *Linguistic Variation and Change*, Oxford: Blackwell.

Ogura, M. 1987. *Historical English Phonology: A Lexical Perspective*, Tokyo: Kenkyusha.

Ogura, M. 1990. *Dynamic Dialectology: A Study of Language in Time and Space.* Tokyo: Kenkyusha.

Ohala, J. 1993. The phonetics of sound change. In: *Historical Linguistics: Problems and Perspectives*, C. Jones (ed.), 237–278, London: Longman.

Pinker, S. 1994. *The Language Instinct* Harmondsworth: Penguin.

Romaine, S. 1982. *Sociohistorical Linguistics*, Cambridge: Cambridge University Press.

Samuels, M.L. 1972. *Linguistic Evolution, with Special Reference to English*, Cambridge: Cambridge University Press.

Smith, J.J. 2000. The letters *s* and *z* in South-Eastern Middle English. *Neuphilologische Mitteilungen* 101: 403–413.

Smith, JJ. 2007. *Sound Change and the History of English.* Oxford: Oxford University Press.

Stuart-Smith, J. 2004. *Phonetics and Philology*, Oxford: Oxford University Press.

Wang, W.S.-Y. 1969. Competing changes as a cause of residue. *Language* 45: 9-–5.

Wells, J. 1982. *Accents of English*, Cambridge: Cambridge University Press.

White, D. 2004. Why we should not believe in short diphthongs. In: *Studies in the History of the English Language II: Unfolding Conversations*, A. Curzan and K. Emmons (eds), 57–84, Berlin and New York: Mouton de Gruyter.

2 Accent as a Social Symbol

LYNDA MUGGLESTONE

Introduction

For Samuel Johnson, drafting his *Dictionary* in the late 1740s, *accent* was already densely polysemous. It could denote patterns of intonation and the prominence given to certain syllables in pronunciation; *antique*, he noted, "was formerly pronounced according to the English analogy, with the accent on the first syllable; but now after the French, with the accent on the last" [my emphases]. By poetic license, *accent* could also signify language or words *per se*. "How many ages hence| Shall this our lofty scene be acted o'er,| In states unborn, and accents yet unknown", states Shakespeare's *Julius Caesar* in an illustrative citation that Johnson included under this sense. In more general terms, *accent*, as Johnson confirms, could indicate "the manner of speaking or pronouncing, with regard either to force or elegance". Supporting evidence from Shakespeare already, however, suggests its potential for qualitative discrimination in this respect, as in the "plain accent" used to describe the forthright speech of Oswald the steward in *King Lear* or Rosalind's "finer" accent in *As You Like It*: "Your accent is something finer than you could purchase in so removed a dwelling." As Puttenham had indicated in his *Arte of English Poesie* (1589), reference models for speech are not to be located in the "ill shapen soundes" of craftsmen or carters or, he adds, "others of the inferiour sort". Even at this point, preference was given to other localized norms, centered on London and surrounding counties within about 40 miles and, in particular, as typified in the usage of educated and courtly speakers –"men ciuill [civil] and graciously behauoured and bred", as Puttenham affirmed.

As Johnson's entry for *accent* suggests, certain meanings are nevertheless prominent only by their absence. Only in the nineteenth century would *accent*, by a process of synecdoche, come to signify the presence of regional marking in speech *per se* – so that one might, or indeed might not, in the idioms of English, "have an accent". "She has a bad figure, she moves ungracefully, perhaps speaks with an accent", an 1865 citation under *accent* in the *Oxford English Dictionary* (*OED*) confirms.

The Handbook of English Pronunciation, First Edition. Edited by Marnie Reed and John M. Levis.

The original definition of *accent* in *OED1*, written in 1884 by the phonetician Alexander Ellis, was telling: "This utterance consists mainly in a prevailing quality of tone, or in a peculiar alteration of pitch, but may include mispronunciation of vowels and consonants, misplacing of stress, and misinflection of a sentence. The locality of a speaker is generally clearly marked by this kind of accent." Illustrative uses include "he has a strong provincial accent" or "an indisputably Irish, Scotch, American … accent".[1] Citational evidence added in the *OED Supplement* (1972), here taken from H.G. Wells's novel *The Autocracy of Mr. Parham* (1930), confirmed the further consolidation of these ideas. "Underbred contradictory people with accents and most preposterous views", wrote Wells, providing an unambiguous correlation between "underbreeding" and "accented" speech. *Underbred*: "Of inferior breeding or upbringing; wanting in polish or refinement; vulgar", the *OED* explains. *Accent*, in Wells's novel, is made to signal the presence of localized marking alongside assumptions that only those lower in the social spectrum will – or should – possess geographical signifiers of this kind. Other evidence added to the *Supplement* (now deleted from *OED3*) made the sociocultural consequences particularly clear: "1956 D. Abercrombie *Prob. & Princ.* iv. 42: *Accent …* is a word which, in its popular use, carries a stigma: speaking *without* an accent is considered preferable to speaking *with* an accent …. The popular, pejorative, use of the word begs an important question by its assumption that an accent is something which is added to, or in some other way distorts, an accepted norm."

The location – both social and linguistic – of Abercrombie's "accepted norm" is equally significant. If "speaking with an accent" had, for Wells, revealed "underbreeding", the opposite end of the social spectrum lay, as White noted in *Words and Their Uses* (1881), in "that tone of voice which indicates breeding". Laden with sociosymbolic values rather different in kind, this form of pronunciation revealed little or nothing of the place of origin of those who used it – whether with reference to what came to be known as "Received Pronunciation" (RP) in Britain, or in the relative homogenization of General American in the United States (see Lippi-Green 1997). As in Abercrombie's analysis, such speakers, in "popular use", were regarded as being able to speak "without an accent" at all. George Bernard Shaw's phonetically-orientated take on the Pygmalion myth in 1914 provides an apt illustration of the sociolinguistic dynamics that can result. Here, the Cockney flower-seller Eliza Doolittle must lose one accent – the geographically marked properties of lower-status London which will, Shaw states, "keep her in the gutter to the end of her days". Courtesy of intensive phonetic re-education, she instead gains another – an "accentless" RP by which, irrespective of social reality, she will pass for a Duchess at the ambassador's garden party. Unlike Cockney, which betokened Eliza's origins – social and regional – in highly specific ways, RP was supra-local, used by speakers "all over the country" as Ellis (1869) had specified, in a speech community characterized by its social meaning as well as its highly restricted membership. As the elocutionist Benjamin Smart (1836) had commented, here with specific reference to accent: "the common standard dialect" is that in which "all marks of a particular place of birth and residence are lost, and nothing appears to indicate any others habits of intercourse than with the well-bred or well-informed, wherever

they may be found." Conversely, it should be remembered that the speech of Northumbrian witnesses, testifying in London in 1861 at the Commission on Mines, was deemed to require an interpreter (Pittock 1997: 118).

While the "received" in other aspects of language practice habitually reflects issues of communality and consensus (see, for example, the early injunction in Cawdrey's *Table Alphabeticall* (1604) "to speak as is commonly receiued"), the history of received pronunciation, and its ideologized values, is instead therefore often bound together with the uncommon or nonrepresentative – the language of the privileged few rather than the accented many. The rise of RP as the prime reference accent can, in this light, seem striking. Examining a range of framing discourses such as education, literature, and the mass media, this chapter will explore the changing role and representation of accent, both localized and supra-local, in the history of English. The patterns of endorsement and emulation which are evident in terms of an emergent RP in, say, the eighteenth-century elocution movement or in the prominence of the supra-local in in the training of announcers on the early BBC (Mugglestone 2008) can, for example, stand in recent years alongside evidence of attitudinal resistance, whether in broadcasting or in the accents one might choose to adopt or shed. Here, too, lexical and semantic shifts provide interesting evidence of change. *Mockney*, a recent entry in *OED3* records, is: "An accent and form of speech affected (esp. by a middle-class speaker) in imitation of cockney or of the speech of Londoners; (generally) *mockney accent*". As in accounts of the British Chancellor George Osborne's attempts at linguistic downshifting (in which traditionally stigmatized features are seen as prominent),[2] a twenty-first century version of *Pygmalion* might well tell a very different story. "People sneered at the chancellor's new mockney accent – but it did make him look more human," wrote Victorian Coren in *The Observer* in April 2013.

Acts of transformation: the eighteenth-century context

Samuel Johnson, it might be noted, steadfastly retained his Staffordshire accent to the end of his days. This, he declared in 1776, was "the purest English". Such patterns of local, and linguistic, allegiance offer a useful corrective to habitual readings by which Johnson is often assumed to be single-handedly standardizing the English of his day.[3] Yet attitudes to Johnson, and his speech, can in fact usefully illuminate a changing consciousness of accent and pronunciation during this period. David Garrick, the famous actor and theatre-manager, who came to London from Lichfield with Johnson in 1735, followed a very different linguistic trajectory. Some eight years younger than Johnson, it is thanks to Garrick's mockery of Johnson's regional marking (a form of speech that Garrick swiftly shed) that we know, for instance, of Johnson's lengthened Staffordshire vowels in words such as *punch*. Rather than commendations of Johnson's accent loyalty, it was perceptions of his "dreadful voice and manner" on which the wife of James Harris, first Earl of Malmesbury (and author of *Hermes*) likewise comments in April 1775.[4] Even James

Boswell's *Life of Johnson* drew attention to Johnson's "uncouth" tones on their first meeting in 1762 (Pottle 2004: 260): "he speaks with a most uncouth voice", Boswell wrote in the intended privacy of his *London Journal*. Of interest too is the diary of Hester Thrale, a close friend of Johnson, who in 1778 decided to award him a score of zero (out of twenty) for "Person and Voice".[5]

The fact that Thrale decided to initiate an evaluative exercise of this kind among her friends is, of course, also significant in this context. Earlier eighteenth-century comment on differences of speech had been decidedly liberal: "A Country Squire … having the Provincial Accent upon his Tongue, which is neither a Fault, not in his Power to remedy", Swift had written, for instance, in 1709. "I do not suppose both these Ways of Pronunciation to be equally proper; but both are used … among Persons of Education and Learning in different parts of the Nation", stated Isaac Watts with similar unconcern (1721: 102). If spelling continued to vary, especially in private use, it clearly also possessed a nationally distributed form; the same was true of the diffusion of a supra-regional grammar. Yet for pronunciation, placed outside the consensus norms of printed texts, there was no public national mode of articulation. The localized, of necessity, remained the norm even if certain modes of pronunciation (e.g., the south-western marking of Somersetshire in Britain) were stereotypically disfavored (see Blank 1996).

The assimilation of accent into regulative discourses of standards and standardization is nevertheless increasingly apparent at this time. Readings of the localized – in the light of what is increasingly promulgated as a supra-regional ideal – can assume strongly negative associations. Boswell himself provides a useful case history. If Boswell is usually remembered in terms of his formative relationship with Johnson, it was in fact Thomas Sheridan, the actor and elocutionist, who was, as Boswell acknowledged, his "Socrates" and mentor. Sheridan's lectures on elocution – emphasizing, in relation to localized language habits, the importance of a wide-ranging shift in attitudes and practice alike – had prompted Boswell's immediate enrolment as Sheridan's private student. "How can consciousness be awoken without information?", Sheridan had declared (1762: 37): "no man can amend a fault of which he is not conscious; and consciousness cannot exert itself when barred up by habit or vanity". Boswell proved a most receptive pupil. "Consciousness" led to repeated anxieties about accent, identity, and regional marking. "Mrs. Miller's abominable Glasgow tongue excruciated me", Boswell wrote in his *London Journal* on March 17, 1762 (Pottle 2004: 221). "Habit" was countered by intentionally corrective "information". Under Sheridan's instruction, Boswell strove to eradicate all traces ("faults") of his Scottish origins from his voice. Similar anxieties later led to an assiduous monitoring of his daughter's speech. If Johnson credited Staffordshire with the "purest English", Boswell did not agree.[6]

In Sheridan's rhetoric, images of "received" speech hence exist alongside a determined inculcation of ideas about what should not be "received" at all. Hitherto, he noted (1762: 37), "many provincials have grown old in the capital, without making any change in their original dialect" (a comment it is tempting to read in the light of Johnson's regionalized speech). In contradistinction, the

regional, for Sheridan, is a firm "mark of disgrace". Placed in the tropes of the "sick" language (an "infection" for which a "cure" is necessary, as Sheridan makes plain), localized speech patterns are framed by the diction of "defect" and "deviation". The accent proposed as the regulative ideal is rather different – not only in its features but also in the perceptual social and cultural values it is made to suggest. It is "a proof that one has kept good company," writes Sheridan, "sought after by all, who wish to be considered as fashionable people, or members of the beau monde" (1762: 30). It is, for Sheridan, an indubitable marker of status or social symbol: "Surely every gentleman will think it worth while to take some pains, to get rid of such evident marks of rusticity," he declares.

Sheridan's "received" speech is both socially and geographically restricted. Prototypically characterizing upper-status speakers in London, it has, as he continues, hitherto "only [been] acquired by conversing with people in polite life". Perry (1775) makes a similar point, selecting "the present practice of polite speakers in London" as his intentionally regulative norm. Nevertheless, as a range of writers indicate, a new democratization of access (and of speech) might henceforth be facilitated through education, elocution, and the national power of print. As Sheridan (1762: 30–31) explained:

> The difficulties to those who endeavour to cure themselves of a provincial or vicious pronunciation are chiefly three. 1st, The want of knowing exactly where the fault lies. 2ndly, Want of method in removing it, and of due application. 3dly, Want of consciousness of their defects in this point.

As we will see, all three were, in a variety of ways, to be provided as the eighteenth and nineteenth centuries advanced. Whereas Johnson's *Dictionary* had merely marked the position of word stress, Sheridan's *Dictionary* (1780) had rather different aims. "One main object … is to establish a plain and permanent standard of pronunciation," the title-page proclaims. Sheridan's work expounds with striking specificity this shift in "consciousness", together with the determined positioning of accent within schema of social meaning. It is nevertheless important to see this as part of a wider process. Buchanan's *Linguae Britannicae vera Pronunciatio* (1757) was, for example, already starting to explore the provision of an "accurate Pronunciation", which native speakers as well as foreigners might acquire by means of lexicography. By 1766, Buchanan had published *An Essay towards Establishing a Standard for an Elegant and Uniform Pronunciation of the English Language … as practiced by the Most Elegant and Polite speakers.* Kenrick's *New Dictionary* (1773) likewise promised full information on "Pronunciation … according to the present practice of polished speakers in the Metropolis". Perry in 1775 made a similar claim. The commodification of accent was also enhanced by the rise of elocution as an industry in a period of marked social change. As an object of desire, the "right accent", characterized by "elegance" rather than "provinciality", might also be acquired, as in Sheridan's teaching of Boswell, or the private lessons offered by a range of other elocutionists across the country (see Benzie 1972).

Pronouncing dictionaries, and other works dedicated to the spoken voice, were disseminated both nationally and internationally,[7] providing an increasingly detailed and prescriptive reference model. This was /h/-full, possessing the velar nasal /ŋ/ rather than /in/ or /iŋg/ in words such as *hopping*, /hw/ rather than /w/ in words such as *which,* using the FOOT-STRUT split, as well as an emergent BATH-TRAP divide. As the elocutionist John Walker (1791: xiii) explained with reference to individual accent modification and the acquisition of "proper pronunciation" (in this instance, the regulative patterning of [v]/ [w]), pronouncing dictionaries were ideally made part of a process of active change:

> Let the pupil select from a dictionary, not only all the words that begin with *v*, but as many as he can of those that have this letter in any other part. Let him be told to bite his under lip while he is sounding the *v* in those words, and to practice this every day till he pronounces the *v* properly at first sight: then, and not till then, let him pursue the same method with the *w*; which he must be directed to pronounce by a putting out of the lips without suffering them to touch the teeth.

Educating accents

"I let other folks talk. I've laid by now, and gev up to the young uns. Ask them as have been to school at Tarley; they've learnt pernouncing; that's come up since my day," comments Mr. Macey in George Eliot's *Silas Marner* (1861). As in the localized metathesis of *pernouncing*, Macey's speech is made to testify to an earlier educational age. Instruction across the late eighteenth and nineteenth centuries, instead increasingly included spoken alongside written language, with a calculated emphasis on the acquisition of supra-regional markers deemed "standard". "It ought to be, indispensably, the care of every teacher of English, not to suffer children to pronounce according to the dialect of that place of the country where they were born or reside, if it happens to be vicious," Buchanan stressed (1757: xli). The potential for social meaning in speech is made particularly explicit: "to avoid a provincial dialect, so unbecoming gentlemen, they are early instructed, while the organs of speech are still flexible, to pronounce properly", Buchanan persuasively declared. Accent, in private education of this kind, is made a telling object of desire.

"Method", as Sheridan had explained, was nevertheless vital. The acquisition of regulative (and supra-local) norms depended in part upon "opening a method, whereby all the children of these realms, whether male or female, may be instructed from the first rudiments, in … the art of reading and speaking with propriety and grace" (1762: 225). This process of acquisition was intended to displace existing practice in which habits of pronunciation "depend entirely upon the common mode of utterance in the several places of [children's] birth and education". Whether by personal tuition (as for Boswell), educational practice in schools and colleges, or conscious application by the motivated individual, the process – and desirability – of educating accents became a prominent topos. The new genre of the pronouncing

dictionary, with its specification of reference models for accent as well as meaning, was presented as particularly useful. The dictionary "must soon be adopted into use by all schools professing to teach English", wrote Sheridan (1762: 261), a precept also evidently taken on board in the emergent national education system in Britain (see Mugglestone 2007: ch.7). "Rp., received pronunciation", as Ellis specified, was "that of pronouncing dictionaries and educated people" (1889: 6).

From the point of view of applied linguistics, elocutionary manuals and educational texts provide considerable detail in this respect. Sheridan's *Elements of English* (1786), aimed at children from the earliest years, provides an obvious example. This sets out detailed guidance by which a "right pronunciation" is to be acquired – and a "wrong" one displaced. The basis of instruction is phonetic, with the order of instruction being first labials, then dentals, labio-dentals, and "palatines". Minimal pairs form the basis of exercises and transcriptions offer disambiguation where necessary, as in the recommended distribution of /ʌ/ or /ʊ/ (*cut*, *bull*) or /hw/ and /w/ (*which*/*witch*) according to supra-regional rather than localized patterns (see, for example, also the specification of rounded [ɒ] after [w] as in *want*, rather than localized [a]). Only favored variants are recorded.

Evidence of the implementation of instruction of this kind is particularly important. Poole's *The Village School Improved* (which had three editions 1813–1815) offers considerable detail of the ways in which, in Enmore in Somerset, children were encouraged to abandon "provincial" forms in favor of supra-local models. Reading aloud became an exercise in discrimination. "Even a coarse or provincial way of pronouncing a word, though sanctioned by the general practice of the district, is immediately noted by the teacher; and exposes the child … as much to the correction of those below him, and consequently to the loss of his place, as any other impropriety in reading would do" (Poole 1815: 40–41). The hierarchical ranking of the class is particularly telling, offering a microcosm of the kind of top-down models of convergence that contemporary works on elocution advocated. Local children, Poole admitted, have habitually "heard and spoken a broad provincial dialect". Learning "to pronounce with propriety" could be challenging: "The more remote the dialect of the [child's] country is from the propriety of the language, the greater is the embarrassment experienced … when he begins to be instructed according to the new and improved system" (1815: 41). Nevertheless, the benefits are presented as incalculable: "this embarrassment is merely temporary" but "permanent advantages are sure to follow", not least in the "intelligent, discriminating manner of reading" and "purity of pronunciation" that will, in the end, be acquired.

Teaching manuals from later in the century provide further evidence of the ways in which reference models of accent were incorporated within general educational practice and assessment. Morrison's *Manual of School Management*, which went through three editions (1859–1863), presents a useful example. Originally "designed for the use of students attending the Glasgow Free Church Training College", the manual sets out recommended methods of instruction on the basis of tried and tested methods. "Nothing has been set down which experience has not proved attainable," Morrison stresses (1863: iii). Exercises within

individual chapters are given as aligned with the Committee of Council of Education "with the view of directing attention to the points considered important by the Inspectors of Schools". An extensive section details "the correct use of letters, the signs of sounds". For the teacher, "the first thing to be done is to analyze the language into its simple elementary sounds"; these again include contrastive medial vowels in *cut* and *bull*, *cat* and *cast*, as well as use of the velar nasal /ŋ/ in words such as *skipping*. As in Sheridan, minimal pairs are advised to enable facility in reading and speaking alike. A section headed "Correct Pronunciation" outlines the principles by which the teaching of reading includes not only comprehension but articulation in the prescribed way: "the first essential requisite in good reading is correct pronunciation" (1863: 125). This, Morrison (1863: 125) points out, is dependent on the teacher suppressing (a) his/ her own regional marking and (b) those of the children in his/her care:

> There is no security that the pupils acquire correct pronunciation, unless the teacher be able to give the example. Accordingly the teacher who is anxious to be in this, as in all things, a model, should strive during his preparatory training to acquire a thorough knowledge of English pronunciation. This can only be done by careful observation of good speakers, or, if need be, by a course of lessons with an accomplished and trustworthy teacher. Whenever the young teacher hears a good speaker pronounce a word differently from what he has been accustomed to, he ought to note it, and never rest satisfied until he has ascertained the correct pronunciation. He will be amazed at the benefit such a course will confer. (1863: 126)

While the teacher's acquisition of "correct orthoepy" is made central to teaching ability in this context, Sheridan's earlier emphasis on "method" is also clear. "The only effectual method by which [the teacher] can secure good pronunciation among his pupils, is to insist that they pronounce every word correctly," writes Morrison: "Constant correction … will alone accomplish the desired result." An educated accent is specified as one devoid of the "peculiarities of pronunciation" which characterize "various districts", whether in terms of "a constant tendency to shorten the long vowels" or "in others to lengthen the short ones", or in the presence other regionally marked features (1863: 126). The normative remit of the teacher is evident: "we advise the teacher, whenever he finds himself located in a particular parish, to observe carefully the prevalent peculiarities; and, when he has done so, vigorously to set himself to correct them among his pupils" (1863: 127). Education reveals, in essence, the firm institutionalization of an ideology in which pronunciation can be divided on standard/subordinate models.

Morrison's strictures are paralleled in a range of other teaching manuals, as well as in school inspectors' reports where articulation (and the absence of regional marking) is often presented as proof of educational success. Recitation – the reading out of a passage with "proper" elocution – was a popular aspect of assessment in which the presence of regional markers could be viewed as testimony not only to local identity but, as other educationalists admonished, as indicators of "Defective Intelligence" *per se*. It was in these terms that John Gill, one of

the most influential writers of teaching manuals in this context (see Hole 2003) chose to orientate his discussion of features such as zero-realization of /h/ or the nonuse of /ʌ/ in *cut*. The classification of purely phonetic features under "Defective Intelligence" amply confirms the negative repercussions of applied language attitudes in educational practice of this kind.

Self-education presents a further domain in which attitudinal shifts to regionally marked speech, and the attempted inculcation of a supra-local model, is in evidence. Texts on pronunciation and elocution often recommended assiduous self-application. It is, however, specific evidence on individual receptiveness to such dictates that can be most illuminating. Prescriptive rhetoric provides merely one side of the story. A useful snapshot here is provided by Michael Faraday, the scientist (and famous lecturer) who began life as the son of a blacksmith in working-class London. It was in this context of self-improvement that Faraday's interest in language, and specifically pronunciation, began. By 1813, he had established, with other members of the local City Philosophical Society, a "mutual improvement plan" whereby some half a dozen friends met "to read together, correct, and improve each other's pronunciation" (see Mugglestone 2011). Five years later, this plan was extended by Faraday's decision to attend Benjamin Smart's lectures on elocution, from which Faraday's detailed notes, running to some 150 pages, remain in the Royal Institution archives in London.

Faraday noted, in full, Smart's maxim: "Always pronounce words according to the best usage of the time … defects or provincialities must be corrected by a dictionary for which purpose I would recommend Walker's or by reference to those who are already correct." Comments on "defective articulation", and its needful remedy, receive equal attention: "H is … the most subject to a corrupt pronunciation and therefore requiring our early attention," Faraday's notebook records; "The person should practice … lists of words beginning with H, then in mixed lists of words some beginning with H, and some with a vowel and lastly with the introduction of the words commencing with H mute." As Smart pointed out, lectures should be accompanied by active practice, not merely passive listening. "Man", Smart added (in another maxim noted down word for word), "is an improving animal … that man only is to be condemned and despised, who is *not* in a state of transition. We are by our nature progressive." Like Sheridan for Boswell, Smart was Faraday's phonetic mentor, in a connection that lasted until the 1850s.

Attitudes, accent, and popular culture

Popular culture also acts as a domain in which the information central to Sheridan's recommended shift in "consciousness" can come into play. The shifts in language practice attested by Boswell and Sheridan, for instance, testify to that process of enregisterment – a cultural awareness of a set of social meanings associated with specific varieties of speech as detailed by Agha (2003, 2005). Cockney, Scots, as well as speech varieties that participate in what Lippi-Green describes as "the myth of non-accent" (1997: 41) all exist, among other varieties, as enregistered

forms across the nineteenth century – and, as Shaw's *Pygmalion* affirms, into the twentieth century too. Literary texts, and the conventions of representation they adopt, can reflect and foster perceptual meanings in this respect with ease.

As in the following extract from George Gissing's *Born in Exile* (1892), conventional orthographical patterning is placed in contrastive distribution with strategic patterns of respelling in the representation of direct speech. Text conventions of this kind rely on acts of reception by which unmodified spelling will, by implication, suggest the standard proprieties of "educated" speech. A social as well as linguistic divide is made to separate Godwin Peake, a student at Whitelaw College, and his uncle; here, a range of approximations denotes the urban vernacular of the London underclass that Godwin's Uncle Joey retains. The textual as well as social asymmetries in representation intentionally encode divisions of identity, education, and status. Yet, as Blake (Austin and Jones 2002: xvii) warns, "Any spelling which differs from th[e] standard may seem bizarre because it is strange; and what is bizarre may often seem ludicrous or comic." Visual disparities of form readily reinforce normative readings of one variety against what can be made to seem unambiguous infelicities and errors in another. Here, stigmatized features such as [Ø] for [h] in '*ow* (*how*), or [in] rather than [iŋ] (*caterin*' against *catering*) are signaled by the inserted apostrophe. As a graphemic marker, this engages with models of deficit rather than difference (indicating the absence of something that "should" be there). Other features (the absence of sandhi phenomena in *a openin*', *a 'int*) are reinforced in intentionally negative readings by their co-occurrence with nonstandard grammar (e.g., *as* relative in "*give a 'int to the young gents as you might come*", alongside multiple negation). The use of socially disfavored lexical items is equally marked. *Gent*, as *OED1* specified in 1899, was "only *vulgar*, exc. as applied derisively to men of the vulgar and pretentious class who are supposed to use the word, and as used in tradesmen's notices".

> '*This* ain't no wye of caterin' for young gents at Collige!' he exclaimed. 'If there ain't a openin' 'ere, then I never see one. Godwin, bo-oy, 'ow much longer'll it be before you're out of you're time over there?'
>
> 'It's uncertain – I can't say.'
>
> 'But ain't it understood as you stay till you've passed the top standard, or whatever it's called?'
>
> 'I really haven't made up my mind what to do.'
>
> 'But you'll be studyin' 'ere for another twelve months, I dessay?'
>
> 'Why do you ask?'
>
> 'Why? cos s'posin' I got 'old o' this 'ere little shop, or another like it close by, me an' you might come to an understandin'—see? It might be worth your while to give a 'int to the young gents as you're in with—eh?'
>
> Godwin was endeavouring to masticate a piece of toast, but it turned to sawdust upon his palate.

Even where pronunciation features are likely to be shared by speakers of different social identities (as in weak forms such as *of* in positions of low stress, or the patterning of ellipsis), they are typically allocated as "accented" and, by implication, "nonstandard" features. Such skewed patterns of representation heighten the assumed contrast between a "standard" – and unmarked – supralocal discourse, against other varieties that are marked, socially and regionally, in a range of ways (see also, for example, American novels and the contrastive marking of accents of the South). Textual patterning of this kind was, by the end of the nineteenth century, a widespread feature of canonical and noncanonical texts alike, appearing in popular journals, newspapers, and magazines, as well as novels.

Factual works can, in fact, be equally productive in the level of language consciousness that they reveal. Entries in the first edition of the *Dictionary of National Biography* (Stephen and Lee 1885–1891) present particularly useful examples, frequently drawing attention to accent as a salient property of identity. "So perfectly fitted was Ainley, both in looks and voice – from which the north country accent had gone during his training under Benson – that he became famous on the first night," we are informed of the actor Henry Ainley; "His short, stout appearance and strong northern Irish accent did not endear him to his contemporaries; Disraeli remarked 'What is that?' on first hearing Biggar speak in the house," the entry for the politician Joseph Biggar states. Entries for Frederick Alexander ("His cultured voice had no trace of regional accent") or Sir Francis Beaufort ("rejected by a school in Cheltenham on the ground that his Irish accent would corrupt the speech of the other boys") share an emphasis on pronunciation as a reference point for social identity. The fact that, in the relatively brief accounts provided, it was seen as important to confirm that William Huskisson had "a most vulgar uneducated accent" or the politician John Felden had a "strong provincial accent" likewise attests to the perceived salience of attitudes of this kind. The *DNB1* entry for the actor Hannah Brand, and the sense of unacceptability her regional accent elicited, is particularly interesting in the light of shifts in language ideology (and recommended changes in praxis) at this time: "Two years later, on 20 March 1794, Brand appeared at the York theatre, playing Lady Townly in Vanbrugh's *The Provoked Husband*. Her manager there, Tate Wilkinson, complained of her old-fashioned dress, provincial accent, conceit, and contradictory passions. All of these provoked the audience, and her performance "met with rude marks of disgustful behaviour".

The broadcast voice

Brand's castigation in terms of accent was intensified because of her prominent position upon the stage – an early model of a broadcast voice. Broadcasting in its modern sense is, of course, a much later phenomenon. In Britain the British Broadcasting Corporation (BBC) – originally the British Broadcasting Company – instituted national radio broadcasting in 1923. Its remit, as its Director General,

John Reith, stressed, was that of public service broadcasting. Ameliorative and beneficial, it was to provide opportunities for access to high culture in what an article in *The Observer* on 18 July 1926 described as a "University of taste". Language was seen as another aspect of such remedial change: "Wireless … can do much to repair … one of the most conspicuous failures of elementary education in raising the quality of common speech." *The Observer* continued: "It could establish – in time – a standard voice analogous to the 'standard yard' and the 'standard pound"' ("Pronunciation Problems" 1926: 17). As Cecil Lewis, an early employee at the BBC, confirmed (1924), "it has often been remarked – and this is one of the responsibilities that are indeed heavy to carry – that the announcing voice sets a fashion in speaking to thousands of homes and should therefore be faultlessly accurate." The ideal, Lewis added, was that of "accentless" speech.

Reith was particularly engaged with the idea of broadcast English as a reference model. Elaborated in his *Broadcast over Britain* (1924), this led to increasingly stringent policies on the kind of accents deemed suitable for announcers. "We are daily establishing in the minds of the public the idea of what correct speech should be and this is an important responsibility," a BBC directive of 1925 specified. As for Sheridan, images of top-down convergence and the need for corresponding emulatory endeavor are marked. As *The Guardian* wrote in December 1932, the BBC's agenda seemed to be that of "levelling up" pronunciation. "You cannot raise social standards without raising speech standards," Arthur Lloyd James, responsible for the training of announcers on the early BBC, had declared. As *The Guardian* reported, "The case for such attempts to level up pronunciation, as put by Mr. Lloyd James, is that it is the business of State education to remove improper, or at any rate socially unpopular, forms of speech behaviour, because this is in practice an obstacle to getting on in the world."[8] If the BBC was, in this, responsive to pre-existing language attitudes, a clearly interventionist remit was also assumed, as Lloyd James (1927) indicates:

> For some reason a man is judged in this country by his language, with the result that there is, broadly speaking, a sort of English that is current among the educated and cultured classes all over the country. It has little local variations, but these are of no matter, and a man who has this sort of accent moves among the rest of his fellow country men without adverse criticism.
>
> This type of speech avoids the lapses of the uneducated and the affectation of the insufficiently educated at both ends of the social scale, and it is the duty of the BBC to provide this sort of speech as often as possible.

While regional speech appeared on local broadcasting, the early BBC effortlessly inculcated the sense of a supra-regional accent as one of its quintessential features, reinforced through accent training in which RP's hegemony was indubitable. That the same practices extended to Australia and Canada (Price 2008), where RP also came to dominate in news broadcasting and announcing, is still more striking.

Belief and behavior: convergence and divergence

Received English, and the acts of reception that surround it, can nevertheless be more complex than the elocutionary rhetoric of Sheridan, Buchanan, or the early BBC can suggest. If responsibility is overtly assumed for the dissemination of one particular "standard" model through the "noble art of printing" by Sheridan or by direct transmission of particular accents (and their associative meanings) on the early BBC, the reality of language practice can, of course, continue to be conspicuously diverse. A supra-regional mode of speech (as Ellis already indicated in the late nineteenth century), RP spans a spectrum of related forms and emerging/obsolescent variants; yod-presence exists alongside yod-absence in words such as *suit* in Ellis's transcribed forms, just as monophthongal variants existed alongside diphthongs in words such as *mate*. Perry's ambition to fix a social model of speech has, in this respect, failed. In Britain, RP is today used by a minority – usually estimated at between 3 and 5% of the population (see, for example, Hughes, Trudgill, and Watt 2012).

Well over 90% of the population has, in these terms, maintained some degree of localized marking in their speech. Accent as a social symbol hence testifies to far more than the indices of the "well-bred", as stressed by Smart, or familiarity with "good company" as Sheridan proclaimed. Outside those accents promoted as "educated" stand, for example, the authority of vernacular culture, of accent loyalty, and of resistance to the ideological hegemonies in which one type of accent alone is favored and the others proscribed. Reactions to the early BBC, and the acts of speech standardization that it attempted to foster, are particularly useful in this context. The privileging of particular forms of speech on the airwaves was not necessarily without resistance. As *The Manchester Guardian* stressed in 1927, "In self-expression we are heretics all, proud of our dialects and our difference." Acknowledging that "the B.B.C. … has attempted to achieve a pact of pronunciation within these islands", it queried whether this could or should be made a shared norm for all. After all, here against the rhetoric of the "accentless", forms of this kind were profoundly "accented" when seen from, say, the perspective of speakers in the Midlands and the North. If RP was supra-regional in use, it remained distinctly southern in its patterning of words such as *fast* and *bath*, *cut* and *bull*. Attempted standardization, the writer continued, was "in many respects a surrender to the slovenly and drawling speech of the Southern English and will be promptly disregarded by all self-respecting speakers of the language" ("Speech control", 1927: 8). Normative readings of accent varieties are not always shared. Images of "disgrace", in Sheridan's terms, can be countered by those of pretension. As in Gaskell's *Mary Barton* (1848), the question of who precisely "talks the right language" can already be made depending on where you are coming from: "'You're frightening them horses,' says he, in his mincing way (for Londoners are mostly all tongue-tied, and can't say their a's and i's properly)", as the Manchester-born John Barton is made to aver.

"Is it wrong for a person to change their accent?", *The Observer* in April 2013 demanded. The social rhetoric it explored exposed the wide-ranging assumptions that have, since the eighteenth century, often informed popular writing on accent. Question of class, social, prejudice, and discrimination all surface in such debates. Since no one accent is inherently better, the arbitrariness of attributions of "disgrace" and "polish" is all too clear. Sheridan's intended democratization in terms of accent now firmly rests in the shared understanding of the perceptual nature of varieties, rather than in pressures for conformity to a top-down ideal. Prestige, too, in this light, is multidimensional. Covert and overt prestige do not pull in the same direction (see, for example, Watson 2006). Specified norms can be rejected; RP, rightly, has been displaced in Australian broadcasting (as well in other domains where national varieties of English now assume pride of place). Like other varieties once promoted as inviolably "correct" (see Lippi-Green 1997), RP is now understood as profoundly accented, not only in its phonological patterning but in the social meanings it has traditionally assumed. Even in news broadcasting on the BBC, it has largely lost its dominance, while transcription policies in *OED3* likewise reflect a commitment to varietal forms. The revised entries of the new *DNB* (Matthew, Harrison, and Goldman 2004) are likewise substantially different in emphasis and orientation. If the supra-local remains a model in language teaching, the hyperlectal features of U-RP (upper-class RP) are not advocated, while notions of the "received" can prompt evident unease. "Because of the dated – and to some people objectionable – social connotations, we shall not normally use the label RP (except consciously to refer to the upper-class speech of the twentieth century)," write Collins and Mees (2003: 3–4). Such shifts of social symbolism are interesting. Alongside the disfavoring of U-RP is, as Coupland and Bishop (2007) confirm, a clear valorization of speakers' own varieties in many (but not all) cases, alongside a decreased responsiveness to supra-local norms in younger speakers. The sociophonetic landscape can nevertheless remain complex. Even in 2013, issues of regional accent and educational delegitimization can still recur. "Cumbrian teacher told to tone down accent," as *The Independent* newspaper stated in November 2013, reporting the views of education inspectors on a school in Berkshire. Alongside the rise of *mockney* and the incorporation of once-stigmatized features such as glottalization within modern RP, the perceptual legacies of the past can linger on.[9]

NOTES

1 The process of revision in *OED3* has now removed the negative coding of Ellis's "mispronunciation … misplacing … misinflection"; see *OED3 accent* sense 7: a. "A way of pronouncing a language that is distinctive to a country, area, social class, or individual". b. "Without possessive or defining word or words: a regional or foreign accent".

2 See, for example, Sam Masters, "George Osborne's 'Man of the People' accent ridiculed", *The Independent* 26 June 2013. http://www.independent.co.uk/news/uk/politics/

george-osbornes-man-of-the-people-accent-ridiculed-8675419.html. Masters isolated Osborne's appropriation of [in] rather than [iŋ], [h]-deletion, and glottalization.

3 Lass's convictions (2000: 57) that, in terms of eighteenth-century phonology, Johnson is a prototypical user of "London standard" are apparently founded on a misapprehension that Johnson hailed from Warwickshire.

4 James Howard Harris (ed.), A Series of *Letters of the First Earl of Malmesbury; His family and Friends from 1745 to 1820* (London: Richard Bentley, 1870), 1: 303.

5 Thrale ranked her friends on a number of factors. See K. Balderston (ed), *Thraliana, the Diary of Mrs. Hester Lynch Thrale (later Mrs. Piozzi) 1776–1809* (Oxford: Clarendon Press, 1942), I.329.

6 See, for example, the comment with which Boswell follows Johnson's linguistic commendation of the regional in Boswell's *Life of Johnson* (1791): "I doubted as to the last article in this eulogy."

7 Five American editions of Perry's dictionary were, for example, published by 1800.

8 The immediate context was the BBC's decision to broadcast a series to schools called "The King's English" in which features such as /h/-dropping and intrusive /r/, as well as a range of regionalized markers, were all proscribed. See "Our London Correspondence", *The Manchester Guardian* 15 December 1932: 8.

9 The robust defence of regional accents, within as well as outside educational contexts, which this event provoked, is, of course, significant in confirming a changing culture of attitudes and praxis in terms of accent in twenty-first century Britain. Equivalent comments in Poole or Morrison by no means elicited censure on the grounds of discrimination or analogies with racism.

REFERENCES

Agha, A. 2003. The social life of cultural value. *Language and Communication* 23: 231–273.

Agha, A. 2005. Voice, footing, enregisterment. *Journal of Linguistic Anthropology* 15: 38-59.

Austin, F. and Jones, B. 2002. *The Language and Craft of William Barnes, English Poet and Philologist (1801–1886)*, Introduction by Norman Blake, Lewiston, NY; Lampeter: Edwin Mellen Press.

Balderston, K. (ed). 1942. *Thraliana, the Diary of Mrs. Hester Lynch Thrale (later Mrs. Piozzi) 1776–1809*, 2 vols., Oxford: Clarendon Press.

Benzie, W. 1972. *The Dublin Orator: Thomas Sheridan's Influence on Eighteenth-Century Rhetoric and Belles Lettres*, Leeds: University of Leeds.

Blank, P. 1996. *Broken English: Dialects and Politics of Language in Renaissance Writings*, New York: Routledge.

Boswell, James. *Boswell's Life of Johnson; Together with Boswell's Journal of a Tour to the Hebrides and Johnson's Diary of a Journey into North Wales*, 6 vols., George Birkbeck Hill (ed.), revised and enlarged by L.F. Powell, 1971, 2nd edition, Oxford: Clarendon Press.

Buchanan, J. 1757. *Linguae Britannicae vera Pronunciatio: or, a New English Dictionary*, London: A. Millar.

Buchanan, J. 1766. *An Essay Towards Establishing a Standard for an Elegant and Uniform Pronunciation of the English Language*, London: Edward and Charles Dilly.

Cawdrey, Robert. 1604. *A Table Alphabeticall Contayning and Teaching the True Writing*

and Vnderstanding of Hard Vsuall English Words. London: Edmund Weauer.
Collins, B. and Mees, I. 2003. *Practical Phonetics and Phonology. A Resource Book for Students*, London: Routledge.
Coren, V. 2013. "George Osborne, gawd bless yer", *The Observer*, 7 April.
Coupland, N. and Bishop, H. 2007. Ideologised values for British accents. *Journal of Sociolinguistics* 11: 74–93.
"Cumbrian teacher told to tone down accent". 2013. *The Independent*, 19 November.
Ellis, A. 1869–1889. *On Early English Pronunciation*, 5 vols., London: Trübner.
Eliot, G. (1861). *Silas Marner: The Weaver of Raveloe*, Edinburgh and London: William Blackwood and Sons.
Gaskell, E. 1848. *Mary Barton. A Tale of Manchester Life*, 2 vols., London: Chapman and Hall.
Gissing, G. 1892. *Born in Exile*, London and Edinburgh: Adam and Charles Black.
Hole, N. 2003. Nineteenth-century method manuals with special reference to English literature. In *Paradigm*, F. Austin and C. Stray (eds.), *The Teaching of English in the Eighteenth and Nineteenth Centuries. Essays for Ian Michael on his 88th Birthday*: 68–75.
Hughes, A., Trudgill, P., and Watt, D. 2012. *English Accents and Dialects. An Introduction to Social and Regional Varieties of English in the British Isles*, 5th edition, London: Edwin Arnold.
Johnson, S. 1755. *A Dictionary of the English Language*, 2 vols., London: W. Strahan.
Kenrick, W. 1773. *A New Dictionary of the English Language*, London: John and Francis Rivington.
Lass, R. 2000. *The Cambridge History of the English Language.* III. *1476–1776*. Cambridge: Cambridge University Press.
Lewis, C. 1924. *Broadcasting from Within*, London: G. Newnes.
Lippi-Green, R. 1997. *English with an Accent: Language, Ideology, and Discrimination in the United States*, London: Routledge.
Lloyd James, A. 1927. "A B.B.C. expert on pronunciation", *Teaching World*, 21 September.
Masters, S. 2013. "George Osborne's 'Man of the People' accent ridiculed", *The Independent* 26/6/2013. http://www.independent.co.uk/news/uk/politics/george-osbornes-man-of-the-people-accent-ridiculed-8675419.html.
Matthew, H., Harrison, B., and Goldman, L. 2004. *Oxford Dictionary of National Biography*, Oxford: Oxford University Press. http://www.oxforddnb.com/.
Morrison, T. 1863. *Manual of School Management*, 3rd edition, Glasgow: William Hamilton.
Mugglestone, L. 2003. Sheridan in the Schoolroom. In *Paradigm*, F. Austin and C. Stray (eds.), *The Teaching of English in the Eighteenth and Nineteenth Centuries. Essays for Ian Michael on his 88th Birthday*: 22–28.
Mugglestone, L. 2007. *Talking Proper. The Rise and Fall of the English Accent as Social Symbol*, Oxford: Oxford University Press.
Mugglestone, L. 2008. BBC English. In the beginning. In *Broadcast English*, J. Schwyter, D. Maillat, and C. Mair (eds.), Arbeiten aus Anglistik und Amerikanistik 33, 197–215, Tübingen: G. Narr.
Mugglestone, L. 2011. Benjamin Smart and Michael Faraday: The principles and practice of talking proper in nineteenth-century England. In *Contours of English and English Language Studies: In Honor of Richard W. Bailey*, M. Adams and A. Curzan (eds.), 87–107, Ann Arbor: University of Michigan Press.
"Our London Correspondence". 1932. *The Manchester Guardian*, 15 December: 8
Perry, W. 1775. *The Royal Standard English Dictionary*, Edinburgh: David Wilson.
Pittock, M. (1997). *Inventing and Resisting Britain: Cultural Identities in Britain and Ireland, 1685–1789*, Basingstoke: Macmillan.
Poole, J. 1815. *The Village School Improved*. 3rd edn. Oxford: Oxford University Press.

Pottle, F. (ed.). 2004. *Boswell's London Journal, 1762–1763*, 2nd edition, Yale: Yale University Press.
Price, J. (2008). New news old news: a sociophonetic study of spoken Australian English in news broadcast speech. In *Broadcast English*, J. Schwyter, D. Maillat, and C. Mair (eds.), Arbeiten aus Anglistik und Amerikanistik *33*, 285–310, Tübingen: G. Narr.
"Pronunciation problems. Further list by the B.B.C." 1926. *The Observer*, 18 July: 17.
Puttenham, G. (1589). *The Arte of English Poesie*, London: Richard Field.
Reith, J. 1924. *Broadcast over Britain*, London: Hodder and Stoughton.
Shaw, B. 1914. *Pygmalion: A Romance in Five Acts*, London: Constable.
Sheridan, T. 1761. *A Dissertation on the Causes of the Difficulties, which Occur in Learning the English Tongue*, London: R. and J. Dodsley.
Sheridan, T. 1762. *A Course of Lectures on Elocution*, London: J. Dodsley.
Sheridan, T. 1780. *A General Dictionary of the English Language*, London: J. Dodsley.
Sheridan, T. 1786. *Elements of English*, London: C. Dilly.
Smart, B. 1836. *Walker Remodelled. A New Critical Pronouncing Dictionary of the English Language*, London: T. Cadell.
"Speech control" (1927). *The Manchester Guardian*, 17 June: 8.
Stephen, L. and Lee, S. (eds.) 1885–1891. *Dictionary of National Biography*, London: Smith, Elder, & Co.
Swift, J. 1709 *A Project for the Advancement of Religion, and the Reformation of Manners*.
Watts, I. 1721. *The Art of Reading and Writing English*, London: John Clark.
Walker, J. 1791. *A Critical Pronouncing Dictionary and Expositor of the English Language*, London: G. Robinson.
Watson, K. 2006. Phonological resistance and innovation in the North-West of England. *English Today* 22: 55–61.
Wells, H.G. 1930. *The Autocracy of Mr. Parham*, New York: Doubleday, Doran and Co.
White, R. (1881). *Words and Their Uses, Past and Present*, New York: Sheldon.

3 History of ESL Pronunciation Teaching

JOHN M. MURPHY AND
AMANDA A. BAKER

Introduction

This chapter tells the story of over 150 years in the teaching of English as a second language (ESL) pronunciation. It is important to acknowledge at the outset that there is little direct evidence of pronunciation teaching practices for most of the modern era of English language teaching (ELT). Prior to the second half of the twentieth century, there were neither video nor audio recordings of pronunciation teachers in action, reflective journaling appears to have been nonexistent (at least not in any retrievable format), and the period's limited number of classroom research reports tended to focus on areas other than pronunciation teaching. Available evidence consists of specialist discussions of language teaching in general and of the teaching of pronunciation. Other sources include several published histories of ELT (e.g., Howatt and Widdowson 2004; Kelly 1969; Richards and Rodgers 2001) and periodic reviews of pronunciation teaching (e.g., Anderson-Hsieh 1989; Leather 1983; Morley 1991, 1994; Pennington and Richards 1986; Pourhosein Gilakjani 2012). Complementing these sources are analyses of English phonology, studies of the acquisition of second language (L2) phonology, teacher training materials, and related research reports. Starting in the 1990s, a few research studies compared the efficacy of different ways of teaching pronunciation (e.g., Couper 2003, 2006; Derwing, Munro, and Wiebe 1997, 1998; Macdonald, Yule, and Powers 1994; Saito 2007; Saito and Lyster 2012a). However, it is only since the early 2000s that researchers have begun to document what typical pronunciation teachers actually do within classrooms (e.g., Baker 2011a, 2011b, 2014), and even these relatively recent contributions include a mere handful of classroom-focused reports.

As valuable as such published sources may be, there is little tangible evidence generated within classrooms of how ESL teachers have been teaching pronunciation

The Handbook of English Pronunciation, First Edition. Edited by Marnie Reed and John M. Levis.

during the past century and a half. One strategy for documenting pronunciation teaching's history, therefore, is to infer from published sources what teachers' likely classroom practices must have been. While traveling this path, it is worth distinguishing between published sources related to classroom events from which classroom practices may be inferred, and the actual classroom behaviors of pronunciation teachers. A close analysis of historical resources may reveal a reliable history of pronunciation teaching. It is also possible, however, that some of the more interesting resources were not all that widely read, assimilated, and applied by classroom teachers. As in many fields, it takes time for specialists' contributions to influence wider audiences.

Before pronunciation teaching (1800–1880s)

A consistent theme within the historical record is that prior to the second half of the nineteenth century pronunciation received little attention in L2 classrooms. While Kelly (1969) reports that 3000 years ago the Sanskrit grammarians of India "had developed a sophisticated system of phonology" (1969: 60) and that educated Greeks of 1800 years ago taught intonation and rhythm to adult learners of Greek, contributions made prior to the nineteenth century were lost over the centuries and failed to influence the modern era. Reflecting ways of teaching Latin to children and young adults of the 1600s–1800s, variations of classical methods, which focused on the rigorous study of grammar and rhetoric, dominated in Europe and the Americas until at least the 1880s (Kelly 1969; Howatt and Widdowson 2004; Richards and Rodgers 2001). Historians group these various methods under the label "the Grammar Translation Method" though a version termed "the Prussian Method" was practised throughout the United States by the mid-1800s (Richards and Rodgers 2001: 5). Teaching methods of the nineteenth century prioritized attention to the written language. While learners were expected to be able to read, understand, and translate literary texts, there was little expectation to speak the language of study. Historians surmise that during this period L2 teachers were not focusing learners' attention on pronunciation at all (see Kelly 1969; Howatt and Widdowson 2004) and for most of the nneteenth century the teaching of pronunciation was "largely irrelevant" (Celce-Murcia et al. 2010: 3).

It is would be a mistake, however, to perceive teaching practices of the 1800s as mere historical curiosities since ways of L2 and foreign language teaching that share much in common with classical methods are widely practised in many parts of the world today (Hu 2005). In China, for example, such a classical approach might be referred to as "the intensive analysis of grammar" while in Korea the label "grammar/reading-based approach" is sometimes used. When pronunciation is taught through such approaches, it typically involves simple repetition of sounds or words (e.g., Baker 2011b). It is also worth keeping in mind that contemporary ways of teaching foreign languages within secondary schools, colleges, and universities throughout the Americas and many other

parts of the world, as noted by Richards and Rodgers (2001), "often reflect Grammar-Translation principles" and that:

> Though it may be true to say that the Grammar-Translation Method is still widely practiced [today], it has no advocates. It is a method for which there is no theory. There is no literature that offers a rationale or justification for it or that attempts to relate it to issues in linguistics, psychology, or educational theory. (Richards and Rogers 2001: 7)

The first wave of pronunciation teaching: precursors (1850s–1880s)

Beginning in the 1850s and continuing for the next 30 years, early innovators such as Berlitz (1882), who was a German immigrant teaching foreign languages in the eastern United States, Gouin (1880) in France, Marcel (1853) in France, and Predergast (1864) in England were rejecting and transitioning away from classical approaches. These specialists in L2 and foreign language teaching were interested in prioritizing speaking abilities, although not necessarily pronunciation specifically. The primary innovation animating their work was to teach learners to converse extemporaneously in the language of study. Such a shift in instructional priorities may seem modest when viewed from a twenty-first century perspective, though their contemporaries would have perceived their proposals as rather odd. The truth is the innovations Marcel, Predergast, and Gouin championed had limited impact within language classrooms of their era, and failed to reach beyond specialist circles (Howatt and Widdowson 2004). This theme of limited impact with respect to specialists' innovations is worth noting since it will recur throughout much of the 150 year period of this review. One of the reasons for lack of impact is that prior to the late 1880s there was no infrastructure (e.g., professional associations, annual conferences, serial publications) through which new ideas about language teaching might have become better known. A consolation is that Marcel, Predergast, and Gouin were academics and their scholarship was known and discussed in specialist circles, especially in Europe. Though their influence in language classrooms was minimal at the time, their scholarship helped set the stage for the emergence of a focus on pronunciation teaching during the next decades. Also, their innovations are reflected in some of the more widely practised language teaching methods of the twentieth century including the Direct (or Natural) Method (e.g., Sauveur 1874), Situational Language Teaching (e.g., Hornby 1950; Palmer 1917), the Natural Approach (Terrell 1977), and the Total Physical Response (Asher 1965).

In contrast to the modest diffusion of Marcel's, Predergast's, and Gouin's innovations, Berlitz developed into a business entrepreneur whose focus on teaching languages for conversational purposes became relatively well known. The first Berlitz language school opened in Providence, Rhode Island, in 1878, with the Berlitz brand reaching its peak of popularity about a quarter century later. By 1914,

the Berlitz franchise had expanded to include 200 language schools throughout England, Germany, and the United States, and as of 2014, there continue to be over 550 Berlitz language schools in at least 70 countries worldwide. For better or worse, Berlitz schools constitute part of the legacy of mid-nineteenth century innovators in language teaching. As Howatt and Widdowson (2004) explain, Berlitz "was not an academic methodologist" but he was "an excellent systematizer of basic language teaching materials organized on 'direct method' lines" (2004: 224). Other than prioritizing the spoken language, most of Berlitz's innovations (e.g., teachers never translate; only the target language is used in the classroom; the teacher is always a native speaker who is supposed to interact enthusiastically with learners) have long been in decline (see Brown 2007). Along with direct and spontaneous use of the spoken language in L2 classrooms, the legacy of the 1850s–1880s innovators includes a style of pronunciation teaching characterized by exposure, imitation, and mimicry. Following Celce-Murcia et al. (2010), we refer to this first wave in the history of pronunciation teaching with the label "*imitative-intuitive*" practice (2010: 2).

The second wave of pronunciation teaching: the reform movement (1880s–early 1900s)

A change that brings us a giant step closer to the modern era, and one that resulted in pronunciation teaching's considerably more consequential second wave, was the formation in Paris during the period 1886–1889 of the International Phonetic Association. Supported by the work of several prominent European phoneticians (e.g., Paul Passy of France, Henry Sweet of England, and Wilhëlm Vietor of Germany), the association formed in response to a societal need to transition away from classical approaches due to advances in transnational travel, migration, and commerce. Passy spearheaded the association's creation, Sweet became known as "the man who taught phonetics to Europe", and Vietor's 1882 pamphlet (initially published in German under a pseudonym) titled *Language Teaching Must Start Afresh!* was both a catalyst for the association's formation and one of the Reform Movement's seminal manifestos. Among the association's earliest and most important contributions was the development circa 1887 of the International Phonetic Alphabet (IPA). Though Passy published the first phonetic alphabet of the modern era in 1888, the International Phonetic Association based what would eventually become known as the International Phonetic Alphabet (IPA) on the work of Sweet (1880–1881). In admiration of this singular accomplishment, Setter and Jenkins (2005) observe that the intention of the IPA's designers was to develop a system of symbols "capable of representing the full inventory of sounds of all known languages" and that its continuing impact on the modern era of pronunciation teaching "is attested by the fact that, over a hundred years later, it is still the universally acknowledged system of phonetic transcription" (2005: 2). In addition to developing the IPA and establishing a scholarly body charged with its continuing revision, the International

Phonetic Association forged interest in pronunciation teaching through promotion of the following core principles (as cited by Celce-Murcia et al. 2010: 3):

- The spoken form of a language is primary and should be taught first.
- The findings of phonetics should be applied in language teaching.
- Teachers must have a solid training in phonetics.
- Learners should be given phonetic training to establish good speech habits.

Although the first principle echoes the innovations of the 1850s–1880s, the next three constitute the association's clearest break with earlier traditions and opened a modern era of pronunciation teaching quite different from the past. Propelled by the convergence of the International Phonetic Association, the four principles the Reform Movement championed, and the development of the IPA, the late 1880s witnessed the first sustained application of *analytic-linguistic* principles to the teaching of pronunciation. The source of the term "analytic-linguistic" to characterize the Reform Movement's continuing impact is the following from Kelly (1969):

> The ways of teaching pronunciation fall into two groups: intuitive and analytical. The first group [i.e., intuitive] depends on unaided imitation of models; the second [i.e., analytic] reinforces this natural ability by explaining to the pupil the phonetic basis of what he [sic] is to do. (1969: 61)

Celce-Murcia et al. (2010) offer a fuller definition of what analytic-linguistic approaches to pronunciation teaching entail. Although their definition reflects the spirit, it probably extends beyond what late noineteenth century reformers originally envisioned:

> An Analytic-Linguistic Approach . . . utilizes information and tools such as a phonetic alphabet, articulatory descriptions, charts of the vocal apparatus, contrastive information, and other aids to supplement listening, imitation, and production. It explicitly informs the learner of and focuses attention on the sounds and rhythms of the target language. This approach was developed [in the late nineteenth century] to complement rather than to replace the Intuitive-Imitative Approach [e.g., Direct Method appeals to mimicry, imitation], aspects of which were typically incorporated into the practice phase of a typical analytic-linguistic language lesson. (Celce-Murcia et al. 2010: 2)

Beginning in the 1890s and continuing throughout the first half of the twentieth century, increasing numbers of language teachers explored and applied the International Phonetic Association's four core principles along with an evolving set of analytic-linguistic instructional techniques for teaching pronunciation. Viewed from a historical perspective, this introduction of analytic-linguistic instructional practices signaled the formation of a "second wave" in the history of ESL pronunciation teaching. The ebb and flow of this second wave would continue for most of the twentieth century. Additional legacies of the International

Phonetic Association are that it established a journal and sponsored regular meetings that were popular with both linguists and language teachers. In effect, as of the 1890s an infrastructure to support the expansion of pronunciation teaching had been born.

Reform movement innovations (1888–1910)

- Findings of phonetics were applied to language teaching and teacher training.
- Formation of pronunciation teaching's second wave through the use of analytic-linguistic instructional techniques.
- The IPA chart served as a classroom tool for teaching pronunciation.
- Instruction focused explicitly on sound segments (consonants and vowels).
- Learners listen to language samples first before seeing written forms.
- In the movement's first decade, teachers tended to provide phonetic information in great detail.
- Later, teachers realized learners could easily become overwhelmed and a focus on phonemic (broader, less detailed) rather than strictly phonetic information became the norm.
- First wave classroom techniques of mimicry and imitation continued; second wave incorporation of phonemic/phonetic information was used to support mimicry and imitation.
- Learners were guided to listen carefully before trying to imitate.
- As one way of practising problematic vowel phonemes, ESL learners might be taught to say quickly and repeatedly two vowel sounds that are near, though not immediately adjacent to, each other on the English phonemic vowel chart. As a practice sequence of rapid repetitions of the two sounds continued the teacher would aim to "harness human laziness" until learners eventually began to produce an intermediate sound located between the two sounds initially introduced (Kelly 1969: 66);
- To raise phonological awareness, ESL students might be asked to pronounce a sentence from their L1 as if a strongly accented native speaker of English were saying it. The intention was to increase learner awareness of pronunciation differences across languages.
- Similarly, to illustrate pronunciation characteristics to be avoided an ESL teacher might pronounce a sentence in English for ESL learners of L1 Spanish backgrounds as if it were spoken by a heavily accented L1 Spanish speaker of English (with Spanish vowels and consonants). Later, the teacher would be able to "refer to this sentence now and again in speaking of the single sounds, as it will serve to warn the students against the kind of mistakes that they themselves are to avoid" (Jespersen 1904: 154)
- Learners were taught to say sentences while mouthing words, consonants, and vowels in an exaggeratedly slow manner. The purpose was to use slow motion speaking as a way of "minimizing interference from the native phonemes and phonological systems" (Kelly 1969: 66);

- For difficulties with consonant clusters in word-final position, an ESL teacher might provide L1 Spanish speakers with practice featuring resyllabification (linking) (i.e., I**t's** a pencil → It –sa pencil; H**e's** a friend → He –sa friend). "As the pupil was made to repeat" such sequences "with increasing speed he [sic] found that he would remake the clusters without inserting the usual Spanish supporting vowel" (Kelly 1969: 67).

Converging and complementary approaches (1890s–1920s)

The emergence of the Reform Movement did not mean that earlier ways of teaching pronunciation were disappearing. In fact, a recurring theme of this review is that two or more orientations toward pronunciation teaching are often in play concurrently. Some teachers work within one orientation or another while others find ways of either synthesizing or moving between different orientations. The coexistence of intuitive-imitative and analytic-linguistic orientations illustrated this phenomenon at the start of the twentieth century. A similar pattern was repeated later in the century with the rise of, for example, the Direct Method, Palmer's Oral Method (1920s), the Audio-Lingual Method and Situational Language Teaching (1960s), Cognitive Code learning (1970s), various designer methods of the 1970s, Communicative Language Teaching (CLT) (1980s), the 1980s–1990s segmental/suprasegmental debate, Task Based Language Teaching (1990s), etc. The pattern is that each orientation introduces an underlying theory, garners specialist attention, prompts the development of teaching practices (and sometimes instructional materials), and informs the work of pronunciation teachers. While different ways of L2 teaching are, as noted by Hyland (2003) in reference to L2 writing instruction, "often treated as historically evolving movements, it would be wrong to see each theory growing out of and replacing the last" (2003: 2). It would be more accurate to describe the different ways of pronunciation teaching witnessed over the past 150 years "as complementary and overlapping perspectives, representing potentially compatible means of understanding the complex reality" of pronunciation teaching (Hyland 2003: 2).

Prior to the initial decades of the Reform Movement (1880s–1890s), the Direct Method had already established roots in the United States and Europe and it continued to gain in popularity well into the twentieth century. Howatt and Widdowson (2004) suggest that the Direct Method probably reached the zenith of its influence in the years leading up to World War I (1914–1918). While Direct Method practitioners (e.g., those working within Berlitz franchise language schools) prioritized the spoken language, they emphasized the intuitive-imitative orientation of pronunciation teaching's first wave and were less interested in providing the degree of explicit phonemic/phonetic information advocated by Reform Movement enthusiasts. Their reticence is understandable since the background of most Direct Method teachers was more likely to have been literature and/or rhetoric rather than the emerging science of phonetics.

The profile of a typical Berlitz teacher of the early twentieth century is also relevant to ELT conditions of the twenty-first century in this regard. Although Berlitz teachers were required to be native speakers of the target language, they were not particularly well trained as either linguists or as teachers beyond short-term workshops provided by the language schools with which they were associated. Howatt and Widdowson (2004) explain that most Berlitz teachers were sojourner adventurer-travelers interested in teaching their native language as a practical means for supporting themselves while seeing the world. As such, this co-occurrence of international enthusiasm for both the Direct Method and the Reform Movement during the initial decades of the twentieth century foreshadows what would be a persistent and continuing theme. As first articulated by Kelly (1969: 61) over 40 years ago, the theme is that intuitive-imitative ways of teaching pronunciation continue to flourish "in the face of competition from [analytic-linguistic] techniques based on phonetics and phonology".

These fundamentally different ways of teaching pronunciation raised two questions: (1) should teachers only ask students to listen carefully and imitate the teacher's pronunciation to the best of their abilities or (2) beyond careful listening and imitating, should the teacher also provide explicit information about phonetics (i.e., how particular features of the sound system operate)? These questions continue to reverberate in contemporary ESL classrooms worldwide. To accomplish the latter was one of the Reform Movement's expressed purposes. Adoption of Reform Movement principles called for a shift in ways of conceiving instructional possibilities by requiring teachers to have specialized training in how the sound system of English operates. Writing a decade after the Reform Movement was well under way but voicing a decidedly pre-1880s perspective, Glauning (1903) suggested that the explicit introduction of information about phonetics "had no place in the classroom, despite the utility of the discipline [of phonetics] to the teacher" (cited in Kelly 1969: 61). In contrast, specialists such as Jesperson (1904) and Breul (1898/1913) believed differently, recommending that "the use of phonetics […] in the teaching of modern languages must be considered one of the most important advances in modern pedagogy, because it ensures both considerable facilitation and an exceedingly large gain in exactness" (Jespersen 1904: 176). As with many present-day ESL teachers, innovators prior to the Reform Movement had not considered possible facilitative effects of providing language learners with explicit information about the sounds and rhythms of the target language. Decades later, many teachers continued (and still continue) to lack sufficient preparation to be able to do so (see Foote, Holtby, and Derwing 2011). While proponents of the Reform Movement were enthusiastic about prioritizing conversational speech, they went further by supporting pronunciation teaching through analytic-linguistic descriptions of, information about, and explicit practice with the sound system being studied. In so doing, they were forming pronunciation teaching's more inclusive second wave, one that embraced both imitative-intuitive and analytic-linguistic ways of teaching pronunciation.

At this point it is important to clarify how the term 'analytic' was used in the early twentieth century since it differs from how the same term is currently applied in contemporary discussions of ESL instructional design (e.g., Long and Crookes 1991). In the context of the Reform Movement the term "analytic" referred to the role of the classroom teacher who had studied the phonological system of the target language, had analyzed its relevant linguistic characteristics in anticipation of classroom teaching, and provided instruction in what the teacher considered to be a manageable number of characteristics through explicit (i.e., deductive, rule-based) instructional procedures. Throughout these various stages, it was the teacher who was responsible for doing the analyzing of the language system while, implicitly, learners were expected to resynthesize (in modern terms) what had been presented to them in order to apply what they were learning to their own pronunciation. The featuring of either an analytic-linguistic component or an even broader analytic-linguistic orientation to pronunciation teaching, along with at least some attention to imitative-intuitive instructional practices, is reflected in most, though not all, of the approaches to pronunciation teaching of the twentieth century and beyond. However, an analytic-linguistic orientation complemented by an integration of both imitative-intuitive and analytic-linguistic instructional practices is featured in most of the more popular pronunciation-dedicated ESL classroom textbooks of the modern era (e.g., Dauer 1993; Gilbert 2012a, 2012b; Grant 2007, 2010).

A period of consolidation (1920s–1950s)

The four decades between the time of the Direct Method's greatest influence (circa 1917) and the heydays of the Audiolingual Method (ALM) in North America and Situational Language Teaching in Great Britain (1960s) offer several lessons. Prior to the 1920s, Reform Movement proponents had already established the importance of understanding how phonological systems operate. Phoneticians interested in English were incredibly productive during this period. Starting early in the 1900s they were documenting its major phonological elements with impressive detail (e.g., Bell 1906; Palmer 1924). By the early 1940s, specialists had provided detailed descriptions of native English speaker (NES) pronunciation including most of its segmental and suprasegmental elements. Kenneth Pike (1945), for example, was an early innovator who provided lasting descriptions of the American English intonation system. Pike's contribution in this area was celebrated by Bolinger (1947: 134) as "the best that has ever been written on the subject" in order to address a need to teach English pronunciation. Pike's identification of a four-point pitch scale (4 = extra high; 3 = high; 2 = mid; 1 = low) has retained its currency, with some of the most prominent teacher guidebooks on pronunciation pedagogy today continuing to use a similar four-point system (e.g., Celce-Murcia et al. 2010). Several years later, linguists in the UK developed similar descriptions of British English intonation (Kingdon 1958a; O'Connor and Arnold 1961) and stress

(Kingdon 1958b), which were regarded as excellent texts for language teachers and learners alike (Pledd 1960; Wells 1998).

By the mid 1950s, Abercrombie had published several innovative discussions of pronunciation teaching (e.g., 1949a, 1949b), which featured prescient discussions of the role of intelligibility and the use of transcription in ESL classrooms (e.g., Abercrombie 1956). It is no exaggeration that Abercrombie's comments on the role of intelligibility, including the need for its prioritization in pronunciation teaching, resonate with contemporary themes (e.g., Brazil 1997; Levis 1999; Munro and Derwing 2011). Of course, specialist descriptions of how the sound system of English operates are continuously being fine-tuned (e.g., Leather 1999; Ladefoged 2006), but most of the basic information about the L1 phonology of English was available by the end of the 1940s. The period 1920s–1950s was a time of consolidation focused on documenting how the sound system of English operated through research into its linguistic code. However, with few notable exceptions (e.g., Clarey and Dixson 1947; Lado, Fries, and Robinett 1954; Prator 1951), less attention was being given to innovations in teaching practices. During the 1920s–1950s specialists were responding to one of the Reform Movement's primary themes: to be able to teach pronunciation language to teachers who need to understand how its phonological system operates.

The decade of the 1930s, a period that was straddled by two world wars, is especially revealing as it coincided with a decline of interest in pronunciation teaching on both sides of the Atlantic. In the United States, the Coleman Report (1929) sparked a national initiative to prioritize the teaching of reading in foreign language classrooms. A similar initiative was also promoted by the British specialist Michael West (e.g., 1927/1935) whose focus on the teaching of reading and vocabulary impacted many parts of the British colonial world. In particular, the Coleman Report proposed "reading first" as an overarching strategy for organizing language instruction along with the principle that development of a reading ability is "the only realistic objective for learners with only a limited amount of study time" (Howatt and Widdowson 2004: 268). Though the Coleman Report focused on the teaching of modern foreign languages and West's recommendations focused on English as a foreign language instruction, their respective influences on the broader field of language education coincided with a period when innovations beyond pronunciation teaching's first two waves were, and would continue to be, curiously missing from the scene.

During this same period, scholars began to question notions of "standard" or "correct pronunciations" of English (Kenyon 1928; McCutcheon 1939; Wilson 1937). With different English dominant countries and diverse regions of those countries having widely varying pronunciations spoken by what was referred to at the time as "cultivated" speakers of English, assumptions that a particular standard of English existed began to decline. As argued by Kenyon (1928: 153),

> …is it so certain as it is so often assumed to be, that uniformity of speech is a supremely desirable end? It certainly is not necessary for intelligibility, for those speakers of the various types of English – Eastern, Southern, and General American, Northern and

> Southern British, and Standard Scottish – who speak their own type with distinctive excellence have no difficulty whatever in understanding one another.

This period, in many ways, represents the origin of more recent trends and advocacy to "teach for intelligibility" among international users of English (e.g., Jenkins 2000). Despite these earlier challenges to standard models of pronunciation, for the rest of the twentieth century descriptions of native English speaker (NES) phonology continued to serve as the basis for "what" to teach in most ESL classrooms worldwide.

Competing conceptual paradigms: 1950–1970s

The 1950s–1970s coincide with a slow rise of attention to innovations in how to teach pronunciation. One way of discerning the instructional practices of a particular era is to examine some of the classroom materials that were available and widely used at the time. This is our strategy for describing some of the innovations during this period. We begin the section by examining four different versions of a text of considerable historical interest titled *Manual of American English Pronunciation* (*MAEP*) (Prator 1951; Prator and Robinett 1957, 1972, 1985). The *MAEP* was a popular ESL course text dedicated to pronunciation teaching used in US colleges and universities as well as other institutions within the US sphere of influence (e.g., Latin America, the Pacific Rim) for well over 20 years. Though its general structure held constant during this period, the *MAEP* was modified several times as its initial author (Clifford Prator) and eventual co-author (Betty Wallace Robinett) continued to expand and revise it through four editions spanning three decades. Differences between its various editions reflect some of the substantive changes in pronunciation teaching between the early 1950s and the mid-1980s. The history of the *MAEP*'s revisions is all the more interesting since its 1951 and 1957 editions preceded the heyday of ALM, while its third and fourth editions came after the field had already begun to experience ALM's decline. Before continuing with a fuller discussion of the *MAEP*, we must first describe the role of pronunciation within ALM to better contextualize pronunciation teaching during the 1960s–1970s, a controversial period of conflicting theoretical perspectives.

ALM and pronunciation teaching (1960–1975): conflicting perspectives

Although the Reform Movement had introduced an analytic-linguistic component to pronunciation teaching decades earlier, classroom procedures well beyond the first half of the twentieth century continued to follow a lesson sequence of information-transmission phases in which a teacher may have

introduced and explained (teachers did not always do so) particular features of English phonology (e.g., sound segments) followed by imitative-intuitive practice opportunities that featured choral and individual repetition, dialogue practice, and other forms of what today would be characterized as teacher-controlled speaking opportunities. As ALM (in the United States) and Situational Language Teaching (in the UK) became widely adopted in the 1960s, imitative-intuitive practice was especially prominent, even if it was occasionally supported by a teacher's analytic-linguistic explanations of phonological features. ALM prioritized attention to spoken forms, though it did so by organizing instruction around oral pattern practice drills and through the intentional overuse (literally) of repetition, mimicry, and memorization. As interest in ALM spread, the tide of pronunciation teaching's first wave (imitative-intuitive) was once again on the rise worldwide. Concurrent advances in technology contributed to the spread of ALM since pattern practice with spoken forms was emphasized both in the classroom and beyond with the support of language laboratories and, a few years later, portable cassette tape players. Spoken accuracy in stress, rhythm, and intonation was prioritized through imitative-intuitive practice, which was right in line with theories of Skinnerian Behavioral Psychology upon which ALM was based. Lamentably, one impact of the heightened international status of ALM during this period was to divert attention away from other innovations in L2 instruction just getting under way, including the Audio-Visual Method in France (e.g., CREDIE 1961), the Council of Europe's Threshold Level project initiative (Van Ek 1973), and Widdowson's (1972) early calls to teach language as communication. At a time when some language instruction specialists were broadening their outlook "and devising new ways of teaching meaning, the [language] lab [as featured in ALM teaching] appeared to be perpetuating some of the worst features of [imitative-intuitive] pattern practice" (Howatt and Widdowson 2004: 319).

Although the "what" of pronunciation teaching had been coming into its own during the 1920s–1960s, the quality of instructional strategies in "how" to teach phonological features stagnated in many classrooms with the rise of ALM. To put it bluntly, ALM's influence led to a suppression of analytic-linguistic innovations as well as a delay in the rise of pronunciation teaching's subsequent waves. On a more positive note, there was a short-lived flirtation with Cognitive Code learning in the early 1970s, a popular theory that described language learning as an active mental process rather than a process of habit formation. Gattegno's (1963) work with the Silent Way in the 1960s–1970s was premised upon similar themes. Some of the implications of Cognitive Code learning might have led to more analytic-linguistic styles of pronunciation teaching but its implications were more often associated with the teaching of grammar. However, the Cognitive Code perspective resonated with at least some teachers' interests in pursuing more analytic-linguistic ways of teaching. Our reason for this brief digression into a discussion of ALM and its impacts during the 1960s and beyond was to set a fuller historical context for the role Prator and Robinett's *MAEP* would play as a precursor to what eventually became pronunciation teaching's "third wave" in the mid-1980s.

Three innovators of the 1960s–1970s: Clifford H. Prator, Betty Wallace Robinett, and J. Donald Bowen

Although Prator and Robinett's *MAEP* is not representative of ALM instructional practices, many of the ESL students of the 1960s–1970s who worked with it had probably completed much of their preceding study of English within ALM-infused classrooms. By the time of its third edition (1972), most ESL teachers were either well aware of ALM instructional practices or were ALM trained themselves. As well as being used in pronunciation-centered ESL courses, the *MAEP* served as a resource for teachers who offered alternative course types (e.g., more broadly focused courses) but who were interested in including some attention to pronunciation. Its 1985 edition coincided with an era of nascent attention to communicative styles of pronunciation teaching, which Prator and Robinett both acknowledged (see 1985: xvi) and attempted to incorporate into the *MAEP*'s final version.

Written with advanced-level ESL student readers in mind, the *MAEP* is filled with well contextualized information on how the sound system of English operates as well as (what were at the times of its various editions) state-of-the-art inventories of controlled and guided practice activities. In a revealing side note, the *MAEP* also supported ESL teacher training within MATESOL/Applied Linguistics courses up until the mid-1980s (Clifford A. Hill, Columbia University, class notes). Since its two earliest editions predated the advents of ALM, Cognitive Code, and CLT, they offer a revealing look into what were some of the more innovative ways of teaching pronunciation during the 1950s–1970s. When viewed from a contemporary vantage point, the *MAEP* illustrates post Reform Movement perspectives, principles, and instructional practices (e.g., explicit attention to phonetic detail, technical explanations, charts, diagrams, as well as additional visual and audio supports). Its several editions were informed by over 60 years of specialist awareness and research into the phonology of English coupled with Reform Movement recommendations on how to teach it. Naturally, the co-authors' original insights played a major role as well. For example, the *MAEP*'s inclusion and sequencing of topics were informed by a needs analysis of "several thousand" international students attending the University of California, Los Angeles (UCLA) over a three-year period (1985: xix). Eventually, the *MAEP*'s 1985 edition incorporated communicative activities with a moderate degree of success (though most would be considered dated by today's standards), an innovation the co-authors discussed as follows:

> The most significant kind of change in the new edition . . . is the result of the effort we have made . . . to introduce more use of language for real communicative purposes in the learning activities for students to carry out. The authors have always shared the belief among teachers that languages cannot really be learned unless they are used for purposes of [genuine] communication. Without communicative intent, pronunciation is not true speech; it is no more than the manipulation of linguistic forms. (1985: xvi)

The *MAEP*'s practice exercises incorporated contextual information and cues to differentiate phonological features including phonemes, thought groups, phonological processes (e.g., linking, assimilations, palatalization, coalescence), suprasegmentals (word stress, sentence stress, rhythm), and intonation (e.g, rising-falling, rising, prominence, affective meaning). Learners were expected to develop a recognition facility in the use of phonemic symbols, and occasionally were asked to transcribe brief segments of speech. Though written for intermediate- to advanced-level ESL readers, its 18 chapters provided learners with extensive technical information on the English phonological system supported with an abundance of practice opportunities. As such, the *MAEP* was a mature illustration of pronunciation teaching's second wave. Even its less successful attempts to incorporate communicative activities illustrate that its authors were anticipating pronunciation teaching's next wave. With the exception of teacher training programs that feature a course dedicated to the teaching of ESL pronunciation, the levels of comprehensiveness and detail about the sound system of English included in the *MAEP* are likely beyond the scope of many ESL teacher preparation courses at the present time (see Burgess and Spencer 2000; Foote et al. 2011; Murphy 1997). The *MAEP*'s decades long publication history illustrates the surprisingly high quality of second wave resources that were starting to be available during the 1950s–1970s. A limitation is that the *MAEP* was designed to be used with relatively advanced-level college and university ESL learners. Though perhaps unintended, an implication was that attention to pronunciation can be delayed until a higher level of language proficiency has been attained by university age ESL learners enrolled in pronunciation-centered courses. This perspective on when and how to focus instruction would be challenged successfully through the contributions of third wave specialists in ESL pronunciation teaching and materials developers of the mid-1980s and beyond.

"Bowen's Technique"

Also active during an era when pronunciation was taught primarily through intuitive-imitative means, Bowen (1972, 1975) developed a novel set of analytic-linguistic techniques for contextualizing pronunciation teaching "with a classic format that is still recommended, for example, by Celce-Murcia and Goodwin (1991) who refer to it as 'Bowen's Technique'" (Morley 1991: 486). Particularly innovative for its time, Bowen (1975) was:

> . . . not a textbook in the usual sense of the term. But a supplementary manual designed to help a motivated student . . . intended to be used along with a [more broadly focused non-pronunciation ESL] text, preferably in short, regular sessions that use only five or ten minutes of the class hour. (Bowen 1975: x)

The teaching strategies central to Bowen's work are described in detail by Celce-Murcia et al. (2010: 9–10 and 147–148). In brief, they involve listening

discrimination and subsequent speaking practice in which minimal pairs are contextualized at the level of whole sentences while supported by the use of visuals, props, physical gestures, and other supports. A core innovation Bowen introduced was to target minimal pair practice beyond the level of individual words by embedding phonological contrasts within whole phrases and sentences. Also, what Bowen defined as a "minimal pair" extended well beyond consonant and vowel phonemes and embraced an ambitious array of phonological processes such as word stress, juncture, prominence, and intonation. Like Prator, Bowen was a second-wave innovator from UCLA who published journal articles and instructional materials during a period when most of his contemporaries were either teaching pronunciation through imitative-intuitive means or were not teaching pronunciation at all. Twenty-four years later Henrichsen, Green, Nishitani, and Bagley (1999) extended the premises of Bowen's work with an ESL classroom textbook and teacher's manual that contextualize pronunciation practice at even broader discourse levels (e.g., whole narratives rather than individual sentences). Chela-Flores (1998) provides another application of Bowen's innovations to the teaching of rhythm patterns of spoken English. In sum, innovators such as Prator, Robinett, and Bowen illustrate that behind the chorus of voices that have been lamenting the demise of ESL pronunciation teaching since the 1970s, there is a fuller backstory to tell.

Designer methods of the 1970s

As reviewed thus far, the professional environment within which ELT takes place has been inconsistent in support for pronunciation teaching. Following ALM's decline in the 1970s, some constituencies (e.g., North American MATESOL programs) seemed preoccupied for a decade or more with what specialists now refer to as the 'designer methods' of the 1970s. Along with ALM and Cognitive Code instructional models as previously discussed, these included Counseling-Learning/Community Language Learning (C-L/CLL), the Silent Way, Suggestopedia, comprehension approaches such as Total Physical Response (TPR) and the Natural Approach, among others. In some cases, their ways of teaching pronunciation contrasted wildly from each other and several were founded on principles reminiscent of debatable values of the past. For example, the unique and poorly understood nature of teacher modeling of the Silent Way depended heavily upon an imitative-intuitive approach, while its proponents argued that they were appealing to learners' analytic abilities to discern linguistic patterns. Suggestopedia might be characterized as an intuitive-imitative approach on steroids since it anticipated students' heightened mental states of 'superlearning' through exposure to massive amounts of scripted spoken discourse. TPR, the Natural Approach, and other comprehension approaches shared the principle that learners should be provided with opportunities to demonstrate comprehension while expectations for learners to begin to speak are delayed. Some of C-L/CLL's explicit purposes that may be of interest were to foster an affectively comfortable classroom, learner-centered

lessons, learner-controlled practice opportunities, as well as analytic-linguistic opportunities to focus on language form (including pronunciation). Eventually, as the field lost interest in designer methods, fewer teachers learned of some of their possibly useful elements (e.g., comprehension approaches' flooding of the learner with well-contextualized spoken input; C-L/CLL's learner-controlled procedure for focusing on pronunciation through use of the "human computer" technique). Following a path charted by Berlitz in the nineteenth century, several of the designer methods became business enterprises, which by the mid-1980s had drifted to the periphery of ESL teaching where they remain today.

The third wave: communicative styles of pronunciation teaching (mid-1980s–1990s)

Along with the final edition of the *MAEP*, the 1980s witnessed CLT's considerable expansion of impact on pronunciation teaching. Emerging from a European tradition, CLT offers a broad orientation to ways of organizing language instruction, which can be applied flexibly depending upon particular contexts of learning and learners' needs. CLT's adaptable nature stands in sharp contrast to the more rigid prescriptions and proscriptions of Berlitz-type orientations as well as the various designer methods of the 1970s. Though CLT principles were well known in specialist circles by the start of the 1980s, it took several more years for methodologists to begin to apply them to ESL pronunciation teaching. Those who did so successfully were ushering in pronunciation teaching's impactful "third wave". In 1983, Marianne Celce-Murcia (also from UCLA) published the first journal article of which we are aware to center on principles and activity-development guidelines for teaching ESL pronunciation through communicative means. Appearing soon afterward, Pica's (1984) journal article featured similar themes. A few years later, Celce-Murcia's (1987) subsequent book chapter followed with an expanded discussion of how to teach pronunciation communicatively. Each of these seminal discussions featured a generous number of activity descriptions illustrating practical ways to implement CLT principles and guidelines as integral dimensions of pronunciation teaching. It is worth noting that both Celce-Murcia and Pica were academic researchers who sometimes served as specialists in ESL instructional methodology. Curiously, the foci of their respective research agendas were areas other than pronunciation teaching. When writing about the teaching of ESL pronunciation they were not reporting empirical studies but were donning the hats of instructional methodologists. There are at least three reasons for proposing that they wore those hats particularly well. Firstly, each of the three publications mentioned was grounded firmly in CLT theory and principles. Secondly, the guidelines presented were easy to understand and remember, even if teachers who lacked training in English phonology may have found them challenging to apply. Thirdly, since the illustration activities Celce-Murcia and Pica provided were straightforward, it was possible for ESL teachers who had requisite background to test them out in their own classrooms.

Celce-Murcia, Pica, and other early third wave innovators of the 1980s (e.g., Acton 1984; De Bot 1983; Gilbert 1978; Morley 1987; Naiman 1987; Wong 1987) had access to professional associations including AAAL, ACTFL, IATEFL, TESOL, and regional affiliates. As a consequence, general CLT themes were already familiar to a growing number of ESL teachers. In contrast to innovators of the 1850s–1880s, by the 1980s a professional infrastructure was in place that featured conventions, serial publications, newsletters, and less formal networking opportunities. Within a few years, Celce-Murcia's (1983, 1987) and Pica's (1984) innovations were being championed by ESL materials developers who would soon publish a succession of innovative pronunciation-centered classroom textbooks.

The third wave's first genre of professional literature: ESL classroom textbooks (mid-1980s–present)

Actually, it is difficult to determine whether or not classroom teachers and materials developers beyond the mid-1980s were directly influenced by innovators such as Celce-Murcia and Pica, or if the impulse to apply CLT principles to the teaching of pronunciation was part of the zeitgeist of the era. Either way, mid-1980s innovations serve as a pivotal historical reference point since ESL methodologists were opening a new path by fusing communicative sensibilities to the imitative-intuitive and analytic-linguistic teaching practices previously established. These innovators inspired three especially useful genres of resource literature, further enhancing pronunciation teaching's third wave. The first genre is textbooks intended to be used in pronunciation-centered ESL courses. Classroom textbooks by Beisbier (1994, 1995), Brazil (1994), Chan (1987), Dauer (1993), Gilbert (1984), and Grant (1993) were organized around CLT principles. They were early examples of third wave classroom textbooks that have continued to grow in number with revised and expanded editions of Gilbert's and Grant's original illustrations (Gilbert 2012b; Grant 2010) along with more recent illustrations such as Cauldwell (2012), Gilbert (2012a), Gorsuch et al. (2012), Grant (2007), Hahn and Dickerson (1999), Hancock (2003), Hewings (2007), Lane (2005), Marks (2007), Miller (2006), and Reed and Michaud (2005).

Of this first genre, Gilbert's *Clear Speech* series (including five separate editions of the original *Clear Speech, Clear Speech from the Start*, and *Speaking Clearly British Edition*) has been the most successful and widely used classroom series focused on teaching ESL pronunciation of the modern era. When asked what were some of the antecedents to her work on the original *Clear Speech* (1984), Gilbert explained:

> Perhaps my earliest influences were Wallace Chafe [1976] who wrote about the prosodic concept of New Information/Old Information and then Joan Morley [1984], who impressed me with the significance of listening comprehension. [Before writing the first *Clear Speech* text] I visited J. Donald Bowen [see above] as he was preparing a draft of *Patterns of English Pronunciation* (1975). From Bowen I adapted the idea of 'minimal sentence pairs,' as opposed to 'minimal word pairs.' This approach led to my most common

form of instructional practice: student pairs give each other a 'minimal sentence pair' choice of answer. If the speaker gets the wrong answer from the listener, then this provides immediate feedback of a conversational breakdown (either in production or listening comprehension). (J. Gilbert, 11/23/2012 personal communication)

The third wave's second genre: activity recipe collections (1990s–2012)

A second genre inspired by mid-1980s innovations is activity recipe collections (ARCs) focused on pronunciation teaching. These are whole books written for ESL teachers that feature descriptions of many dozens of pronunciation activity prototypes. The fact that the three earliest illustrations of the genre (Bowen and Marks 1992; Hancock 1996; Laroy 1995) were written by British specialists may be a reflection of CLT's European roots. Their books differ from first genre teaching materials since ARCs are not classroom textbooks. Rather, ARCs are book-length collections of stand-alone activities designed as resources for teachers to digest, tailor to their own contexts of teaching, try out in ESL classrooms, and modify as needed. While ARCs had previously been established as a teacher resource staple of the field for the teaching of grammar, reading, spoken fluency, and writing (e.g., Hedge 1988; Ur 1988), Bowen and Marks (1992) is the first ARC dedicated to communicative ways of teaching pronunciation while Hewings (2004) and Brown (2012) are the genre's most recent illustrations. With the exception of the latter, as well as short sections of Bailey and Savage (see 1994: 199–262) and Nunan and Miller (see 1995: 120–150), those currently available feature British styles of pronunciation.

The third wave's third genre: teacher preparation texts (late 1990s–present)

The final decades of the twentieth century witnessed another notable advance and with it a third genre of professional literature: the publication of high-quality resource books dedicated to the preparation of ESL pronunciation teachers. As of 2014, over a dozen examples of this genre have been published, most notably Celce-Murcia, Brinton, and Goodwin (1996) (followed by a 2010 revised and expanded edition), Lane (2010), and Rogerson-Revell (2011). While Celce-Murcia et al. and Lane prioritize patterns of North American pronunciation, Rogerson-Revell's is a specifically British text. In contrast, Walker (2010) focuses not on teaching traditional native speaker standards of English pronunciation but the pronunciation of English as a Lingua Franca (ELF). Kenworthy (1987) merits special attention since it was the first teacher preparation volume of the modern era to focus on how to teach ESL pronunciation. Also, its publication coincided with the centennial anniversary of the birth of the Reform Movement. Other notable examples include Avery and Ehrlich (1992), Dalton and Seidlhofer (1994), Underhill (1994), Fraser (2001), Gilbert (2008), Kelly (2000), Lane (2010), as well as an early

booklet by Wong (1987) and later booklets by Murphy (2013) and Poedjosoedarmo (2003). A central feature each of these texts shares is their sustained focus on how to teach ESL pronunciation, a focus Burgess and Spencer (2000), Burns (2006), Foote, Holtby, and Derwing (2011), and Murphy (1997) document as lacking in many contemporary ESL teacher preparation programs. Availability of this very helpful genre of teacher preparation material is fitting testimony to the efforts of pronunciation teaching specialists of the preceding 150 years.

Pronunciation teaching specialists (1980s–1990s)

In addition to inspiring three new genres of published resources to support ESL pronunciation teaching, third wave innovators of the mid-1980s also prompted a trend in the type of specialist who would drive the field of pronunciation teaching for the next two decades. The trend was that during the 1980s–1990s the most influential authors and conference presenters on the topic of pronunciation teaching were specialists in instructional methodology (e.g., William Acton, Donna Brinton, Berta Chela-Flores, Wayne Dickerson, Suzzane Firth, Judy Gilbert, Janet Goodwin, Joanne Kenworthy, David Mendelsohn, John Levis, Joan Morley, John Murphy, Neil Naiman, Charles Parish, Martha Pennington, Jack Richards, Earl Stevick, and Rita Wong) and/or materials developers (e.g., Tim Bowen, Rebecca Dauer, Judy Gilbert, Carolyn Graham, Linda Grant, Mark Hancock, Lynn Henrichsen, Martin Hewings, Linda Lane, Clement Laroy, Jonathan Marks, Sue Miller, and Gertrude Orion). Though prominent in the field, these specialists tended not to be empirical researchers, at least not in connection with the teaching of pronunciation. Echoing the models of Celce-Murcia and Pica a decade earlier, some had research agendas focused on areas other than pronunciation teaching. However, a theme worth highlighting is that pronunciation specialists of the 1980s–1990s were not conducting empirical investigations on topics such as which dimensions of L2 phonology are more important to teach or how they might be most effectively taught in language classrooms. For the most part, they were basing their recommendations for pronunciation teaching on (a) their own familiarity with relevant literatures (i.e., they were reading widely and synthesizing well), (b) their experiences as teachers of pronunciation, and (c) their intuitions. While the research base may have been thin, third wave specialists of the 1980s–1990s were successful in integrating imitative-intuitive, analytic-linguistic, and communicative means of teaching pronunciation.

Ontogeny of ESL pronunciation teaching in the twentieth century

Implicit in the published work of specialists and materials developers of the 1980s–1990s were provisional answers to some essential research questions (e.g., Which features of English phonology are more important to teach? What

is the best sequence for teaching them? Which teaching strategies and methods of teaching are most effective?), but there remained little in the way of empirical research to support their work. This lack of relevant research may reflect the degree of maturation in the field of ESL pronunciation teaching at the time. Nearly a century before, the Reform Movement had given birth to the modern era by establishing pronunciation teaching as a reputable endeavor and introducing an analytic-linguistic perspective on how to teach. The initial decades of the twentieth century witnessed a period of the field's early childhood as research documentation grew concerning how the sound system of English operates along with concurrent blending of both imitative-intuitive and analytic-linguistic instructional approaches. The mid-twentieth century coincided with a period perhaps best characterized as pronunciation teaching's adolescence. There were early efforts to increase the proportion of analytic-linguistic ways of teaching along with tentative efforts to introduce communicative themes. However, we can also see that advances in pronunciation teaching experienced a maturational backslide in the 1960s as ALM prioritized the imitative-intuitive orientation at the expense of what might have been more substantive innovations. In many parts of the world this stagnation continued throughout the 1970s as confusion continued over how to respond to the wider field's embrace of CLT. Another condition that siphoned attention away from pronunciation teaching during the 1970s–1980s was growing interest in the teaching of L2 reading and L2 writing, a period when ESL learners faced considerable academic literacy demands. L2 reading and L2 writing scholarship was at center stage for ESL teachers who completed their professional training throughout the 1980s–1990s. While L2 pronunciation research lagged behind, L2 reading and L2 writing researchers became some of the field's most prominent leaders. The generation of teachers and scholars they trained comprise a large proportion of today's ESL teachers, material developers, teacher educators, and researchers. Some of the impacts of this historical course of events continue to be felt today. For over two decades, for example, we have had access to a highly respected journal dedicated specifically to L2 writing and to several even more established journals in which L2 reading research dominates. However, a journal dedicated to L2 pronunciation, the *Journal of Second Language Pronunciation*, is scheduled to appear for the first time in 2015. The closest comparable serial publication currently available is *Speak Out!*, a newsletter of IATEFL's Pronunciation Special Interest Group. As often happens with young adults, the teaching of ESL pronunciation from the 1960s through the early 1980s was experiencing a phase of uncertainty and indecision. By the mid 1980s, however, third wave methodologists had begun to explore a more mature direction of instructional possibilities. In the 1990s, this direction was embraced by an even larger number of specialist writers and materials developers. Fortunately, the quality of their work would be further enhanced near the start of the twenty-first century as empirical researchers began to address a series of unresolved research topics.

A gap in ESL pronunciation teaching (up until the mid-1990s)

Along with the many advances witnessed through the three waves of instructional innovations described thus far, specialists were not producing primary empirical research that advanced the quality of pronunciation teaching. Evidence of this lack of empirical research support may be found in Brown's (1991) then state-of-the-art edited collection. Though one chapter is grounded in empirical research (Brown's own discussion of functional load), the collection included no other such examples. As Deng et al. (2009) point out, Brown (1991) lamented in his introduction that "second language pronunciation research did not receive the degree of attention it merited from researchers" (1991: 1). Eighteen years later, Deng et al. (2009) reviewed 14 top tier Applied Linguistics journals for the period 1999–2008 and found that "pronunciation is still underrepresented in the [professional research] literature" (2009: 3). It would not be until the mid-1990s that the work of a small number of empiricists began to fill the gap Brown (1991) and Deng et al. (2009) identified. Research studies by Macdonald, Yule, and Powers (1994), Munro and Derwing (1995), and Wennerstrom (1994) initiated a modern era of primary empirical research to inform the work of ESL pronunciation teaching, an era constituting the field's contemporary 'fourth wave.'

The fourth wave: emergence of empirical research (mid-1990s–present)

A final theme offered as a way of closing this review reflects recent empirical research being used to inform the teaching of ESL pronunciation. It took well over a century for the Reform Movement to culminate in the growing number of fourth wave empirical researchers who are now investigating topics in three macro-level areas of focus: (1) what features of ESL phonology are necessary to teach; (2) how to effectively teach them, and (3) what teachers and students believe and know about pronunciation instruction. Though there is insufficient space to do justice to all that has been published since the mid-1990s, a few representative examples are provided below in Table 3.1. The studies are categorized according to macro-level themes that relate most closely to one of the three topic areas posed above. The majority of the studies listed under the table's first two macro-level themes represent experimental or quasi-experimental investigations that are at least partially connected to the teaching of ESL pronunciation. In addition, a number of researchers have recently begun to explore some of the dynamic connections that exist between teachers' and students' beliefs and actual (or reported) classroom practices. This most recent research agenda is represented in the table's final section, focusing on teachers' cognition (knowledge and beliefs) and learners' perception about pronunciation instruction. Considered collectively, the three sections constitute the heart of the fourth wave of pronunciation teaching and illustrate several research agendas for the future.

Table 3.1 Empirical research that supports ESL pronunciation teaching (ESL pronunciation teaching's fourth wave).

Theme	*Empirical studies (examples)*
Macro-level Theme A: exploring what to teach in English pronunciation	
Theme 1: • Effects of segmentals and supra-segmentals on the intelligibility/comprehensibility of L2 speech and implications for teaching ESL	• Field (2005) • Hahn (2004) • Llurda (2000) • Munro and Derwing (1995, 1998) • Trofimovich and Baker (2006) • Zielinski (2008)
Theme 2: • Effects of sociocultural factors on the intelligibility/comprehensibility of L2 speech and implications for teaching ESL	• Bent and Bradlow (2003) • Deterding (2005) • Deterding and Kirkpatrick (2006) • Kang (2012) • Kennedy and Trofimovich (2008) • Matsuura (2007) • Munro, Derwing, and Morton (2006) • Trofimovich and Baker (2006)
Theme 3: • Contrasting analyses of L1 and L2 English speakers' production and implications for teaching ESL	• Low (2006) • Pickering (2001, 2004) • Pickering, Hu, and Baker (2012) • Setter (2006) • Wennerstrom (1994)
Macro-level Theme B: exploring how to teach pronunciation effectively	
Theme 1: • Establishing priorities in pronunciation instruction	• Derwing, Munro and Wiebe (1998) • Jenkins (2000) • Munro and Derwing (2006) • Saito (2011)
Theme 2: • Impact of instruction and/or feedback on learner intelligibility and/or phonological improvement	• Couper (2003, 2006, 2011) • Derwing, Munro, and Wiebe (1997) • Dlaska and Krekeler (2013) • Levis and Pickering (2004) • Lord (2008) • Macdonald, Yule, and Powers (1994) • Saito (2007) • Saito and Lyster (2012a) • Tanner and Landon (2009) • Trofimovich, Lightbown, Halter, and Song (2009) • Trofimovich and Gatbonton (2006)
Theme 3: • Pronunciation strategies for successful oral communication	• Osburne (2003)

Continued

Table 3.1 (*Cont'd*)

Theme	*Empirical studies (examples)*
Macro-level Theme C: teachers' cognitions (beliefs and knowledge) and learners' perspectives on pronunciation instruction	
Theme 1: • Learners' preferences regarding pronunciation instruction, feedback and accents	• Kang (2010) • Scales, Wennerstrom, Richard, Wu (2006)
Theme 2: • Learners' language awareness, aural comprehension skills and improved pronunciation	• Kennedy and Trofimovich (2010) • Saito (2013)
Theme 3: • Teachers' beliefs and knowledge about pronunciation instruction	• Baker (2011a, 2011b, 2014) • Foote, Holtby, and Derwing (2011) • Jenkins (2005) • Macdonald (2002) • Saito and Lyster (2012b) • Sifakis and Sougari (2005)

Finally, if we may speculate on the future of ESL pronunciation teaching, there is every reason to feel optimistic. Having completed this historical review, we sense a momentum building, which suggests that a fifth wave of innovations is likely to appear within the coming decade. Along with continued synthesis of the four waves identified thus far (i.e., imitative-intuitive, analytic-linguistic, and communicative ways of teaching, along with the development of an empirical research base to support instructional innovations), we believe that the eventual infusion of empirical research findings in materials development, teacher training, and teachers' actual classroom practices will serve to constitute pronunciation teaching's next (i.e., fifth) wave.

REFERENCES

Abercrombie, D. 1949a. Teaching pronunciation. *English Language Teaching* 3(5): 113–122.

Abercrombie, D. 1949b. Some first principles: Part 1. English Language Teaching 3(6): 141–146; Part 2. *English Language Teaching* 3(7): 169–171.

Abercrombie, D. 1956. *Problems and Principles: Studies in the Teaching of English as a Second Language*, London: Longmans, Green.

Acton, W. 1984. Changing fossilized pronunciation. *TESOL Quarterly* 18: 71–85.

Anderson-Hsieh, J. 1989. Approaches toward teaching pronunciation: a brief history. *Cross-Currents* 16: 73–78.

Asher, J. 1965. The strategy of the total physical response: an application to learning Russian. *International Review of Applied Linguistics* 3 291–300.

Avery, P. and Ehrlich, S. 1992. *Teaching American English Pronunciation*, Oxford: Oxford University Press.

Baker, A.A. 2011a. Discourse prosody and teachers' stated beliefs and practices. *TESOL Journal* 2 263–292.

Baker, A.A. 2011b. ESL teachers and pronunciation pedagogy: exploring the development of teachers' cognitions and classroom practices. In: *Proceedings of the 2nd Pronunciation in Second Language Learning and Teaching Conference*, J.M. Levis and K. LeVelle (eds.), September 2010, 82–94, Ames, IA: Iowa State University.

Baker, A.A. 2014. Exploring teachers' knowledge of L2 pronunciation techniques: teacher cognitions, observed classroom practices and student perceptions. *TESOL Quarterly, 48*, 136–163. doi: 10.1002/tesq.99.

Bailey, K. M. and Savage, L. 1994. *New Ways in Teaching Speaking*, Alexandria, VA: TESOL.

Beisbier, B. 1994. *Sounds Great: Low-Intermediate Pronunciation for Speakers of English*, Boston: Heinle and Heinle Publishers.

Beisbier, B. 1995. *Sounds Great: Intermediate Pronunciation for Speakers of English*, Boston: Heinle and Heinle Publishers.

Bell, A.G. 1906. *The Mechanism of Speech*, New York: Funk & Wagnalls.

Bent, T. and Bradlow, A. 2003. The interlanguage speech intelligibility benefit. *Journal of the Acoustical Society of America* 114: 1600–1610.

Berlitz, M. 1882. *Méthode pour L'enseigement la Langue Française dan les Ecoles Berlitz, Première Partie*, Boston, MA: Schoenhof (with E. Dubois).

Bolinger, D.L. 1947. American English intonation. *American Speech* 22(2): 134–136.

Bowen, D. 1972. Contextualizing pronunciation practice in the ESOL classroom. *TESOL Quarterly* 6: 83–94.

Bowen, D. 1975. *Patterns of English Pronunciation*, Rowley, MA: Newbury House.

Bowen, T. and Marks, J. 1992. *The Pronunciation Book: Student-Centered Activities for Pronunciation Work.*, New York: Longman.

Brazil, D. 1994. *Pronunciation for Advanced Learners of English*. New York: Cambridge University Press.

Brazil, D. 1997. *The Communicative Value of Intonation in English*. New York: Cambridge University Press.

Breul, K. 1898/1913. *The Teaching of Modern Foreign Languages and the Training of Teachers*. New York: Cambridge University Press.

Brown, A. (ed.). 1991. *Teaching English Pronunciation: A Book of Readings*,New York: Routledge.

Brown, H.D. 2007. *Teaching by Principles: An Interactive Approach to Language Pedagogy*, 3rd edition), Englewood Cliffs, NJ: Prentice-Hall Regents.

Brown, J.D. 2012. *New Ways in Teaching Connected Speech*, Alexandria, VA: TESOL International Association.

Burgess, J. and Spencer, S. 2000. Phonology and pronunciation in integrated language teaching and teacher education. *System* 28: 191–215.

Burns, A. 2006. Integrating research and professional development on pronunciation teaching in a national adult ESL program. *TESL Reporter 39*,: 34–41.

Cauldwell, R. 2012. *Cool Speech: Hot Listening, Cool Pronunciation* [iPad application], Birmingham, UK: Speech in Action.

Celce-Murcia, M. 1983. Activities for teaching pronunciation communicatively. *CATESOL News* May: 10–11.

Celce-Murcia, M. 1987. Teaching pronunciation as communication. In: *Current Perspectives on Pronunciation: Practices Anchored in Theory*, J. Morley (ed), 5–12, Washington, DC: TESOL.

Celce-Murcia, M., Brinton, D.M., and Goodwin, J.M. 1996. *Teaching Pronunciation: A reference for teachers of English to speakers of other languages*. Cambridge: Cambridge University Press.

Celce-Murcia, M. and Goodwin, J. 1991. Teaching pronunciation. In: *Teaching English as a Second or Foreign Language*, 2nd edition M. Celce-Murcia (ed.), 136–153, New York: Newbury House.

Celce-Murcia, M., Brinton, D.M., Goodwin, J.M., and Griner, B. 2010. *Teaching Pronunciation: A Reference for Teachers of English to Speakers of Other Languages*, 2nd edition, Cambridge: Cambridge University Press.

Chafe, W. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In: *Subject and Topic*, C.N. Li (ed.), 25–55, New York: Academic Press.

Chan, M. 1987. *Phrase by Phrase: Pronunciation and Listening in American English*, Englewood Cliffs, NJ: Prentice Hall.

Chela-Flores, B. 1998. *Teaching English Rhythm: From Theory to Practice*, Caracas, Venezuela: Fondo Editorial Tropykos.

Clarey, E. and Dixson, R.J. 1947. *Pronunciation Exercises in English*, New York: Regents.

Coleman, A. 1929. *The Teaching of Modern Foreign Languages in the United States*, vol. 12, New York: American and Canadian Committees on Modern Languages.

Couper, G. 2003. The value of an explicit pronunciation syllabus in ESOL teaching. *Prospect* 18: 53–70.

Couper, G. 2006. The short and long-term effects of pronunciation instruction. *Prospect* 21: 46–66.

Couper, G. 2011. What makes pronunciation teaching work? Testing the effect of two variables: socially constructed metalanguage and critical listening. *Language Awareness* 20(3): 159–182. doi: 10.1080/09658416.2011.570347.

CREDIE. 1961. *Voix et Images de France: Cours Audio-Visuel de Francais Premier Degree*, London: Harrap. Originally published St Cloud, 1958.

Dalton, C. and Seidlhofer, B. 1994. *Pronunciation*, Oxford: Oxford University Press.

Dauer, R. 1993. *Accurate English: A Complete Course in Pronunciation*. Englewood Cliffs, NJ: Prentice-Hall Regents.

De Bot, K. 1983. Visual feedback on intonation I: Effectiveness and induced practice behavior. *Language and Speech* 26: 331–350.

Deng, J., Holtby, A., Howden-Weaver, L., Nessim, L., Nicholas, B., Nickle, K., Pannekoek, C., Stephan, S., and Sun, M. 2009. *English Pronunciation Research: The Neglected Orphan of Second Language Acquisition Studies? (WP 05-09)*, Edmonton, AB: Prairie Metropolis Centre.

Derwing, T.M., Munro, M.J., and Wiebe, G. 1997. Pronunciation instruction for "fossilized" learners: Can it help? *Applied Language Learning* 8: 217–235.

Derwing, T.M., Munro, M. ., and Wiebe, G. 1998. Evidence in favor of a broad framework for pronunciation instruction. *Language Learning* 48: 393–410.

Deterding, D. 2005. Listening to estuary English in Singapore. *TESOL Quarterly* 39(3): 425–440.

Deterding, D. and Kirkpatrick, A. 2006. Emerging South-East Asian Englishes and intelligibility. *World Englishes* 25(3/4): 391–409.

Dlaska, A. and Krekeler, C. 2013. The short-term effects of individual corrective feedback on L2 pronunciation. *System* 41(1): 25–37. doi: http://dx.doi.org/10.1016/j.system.2013.01.005.

Field, J. 2005. Intelligibility and the listener: the role of lexical stress. *TESOL Quarterly* 39: 399–423.

Foote, J.A., Holtby, A.K., and Derwing, T.M. 2011. Survey of the teaching of pronunciation in adult ESL programs in Canada, 2010. *TESL Canada Journal* 29: 1–22.

Fraser, H. 2001. *Teaching Pronunciation: A Handbook for Teachers and Trainers: Three Frameworks for an Integrated Approach*, Sydney, Australia: TAFE NSW Access Division.

Gettegno, C. 1963. *Teaching Foreign Languages in Schools: The Silent Way*, Reading, England: Educational Explorers Ltd.

Gilbert, J.B. 1978. Gadgets: non-verbal tools for teaching pronunciation. *CATESOL Occasional Papers* 4: 68-78.

Gilbert, J.B. 2008. Teaching pronunciation: using the prosody pyramid. Retrieved from: http://www.cambridge.org/other_files/downloads/esl/booklets/Gilbert-Teaching-Pronunciation.pdf.

Gilbert, J.B. 2012a. *Clear Speech from the Start: Basic Pronunciation and Listening Comprehension in North American English*, 2nd edition, New York: Cambridge University Press.

Gilbert, J.B. 2012b. *Clear Speech: Pronunciation and Listening Comprehension in North American English*, 4th edition (1st edition, 1984), New York: Cambridge University Press.

Glauning, F. 1903. *Didaktik und Methodik des englischen Unterrichts*, Munich.

Gorsuch, G., Meyers, C., Pickering, L., and Griffee, D. 2012. *English Communication for International Teaching Assistants*, 2nd edition, Long Grove, IL: Waveland Press, Inc.

Gouin, J. 1880. *The Art of Studying and Teaching Languages*, Translation by H. Swan and V. Bétis. Originally published in Paris, 1880, London: George Philip.

Grant, L. 1993. *Well Said: Advanced English Pronunciation*, Boston, MA: Heinle & Heinle.

Grant, L. 2007. *Well Said Intro: Pronunciation for Clear Communication*, Boston: Thomson/Heinle & Heinle.

Grant, L. 2010. *Well Said: Pronunciation for Clear Communication*, 3rd edition (1st edition, 1993), Boston, MA: Thomson/Heinle & Heinle.

Hahn, L.D. 2004. Primary stress and intelligibility: research to motivate the teaching of suprasegmentals. *TESOL Quarterly* 38: 201–223.

Hahn, L. and Dickerson, W. 1999. *Speech Craft: Discourse Pronunciation for Advanced Learners.* Ann Arbor, MI: University of Michigan Press.

Hancock, M. 1996. *Pronunciation Games*, Cambridge: Cambridge University Press.

Hancock, M. 2003. *English Pronunciation in Use* (intermediate), Cambridge: Cambridge University Press.

Hedge, T. 1988. *Writing: Resource Books for Teachers*, New York: Oxford University Press.

Henrichsen, L., Green, A., Nishitani, A., and Bagley, C.L. 1999. *Pronunciation Matters: Communicative, Story-based Activities for Mastering the Sounds of North American English*, Ann Arbor, MI: University of Michigan.

Hewings, M. 2004. *Pronunciation Practice Activities: A Resource Book for Teaching English Pronunciation*, Cambridge: Cambridge University Press.

Hewings, M. 2007. *English Pronunciation in Use* (advanced), Cambridge: Cambridge University Press.

Hornby, A.S. 1950. The situational approach to language teaching (I), (II), and (III). A series of six articles. *English Language Teaching Journal* 4(4); 98–103; 4(5): 121–128; and 4(6): 150–156.

Howatt, A.P.R. and Widdowson, H.G. 2004. *A History of English Language Teaching*, 2nd edition, New York: Oxford University Press.

Hu, G. 2005. Contextual influences on instructional practices: a Chinese case for an ecological approach to ELT. *TESOL Quarterly* 39: 635–660.

Hyland, K. 2003. Writing and teaching writing. In: *Second Language Writing*, New York: Cambridge University Press.

Jenkins, J. 2000. *The Phonology of English as an International Language*, Oxford: Oxford University Press.

Jenkins, J. 2005. Implementing an international approach to English pronunciation: the role of teacher attitudes and identity. *TESOL Quarterly* 39: 535–543.

Jespersen, O. 1904. *How to Teach a Foreign Language*,S.Y.-O. Bertelsen (translator), London: George Allen & Unwin. http://archive.org/details/howtoteachforeig00jespuoft.

Kang, O. 2010. ESL learners' attitudes toward pronunciation instruction. In: *Proceedings of the 1st Pronunciation in Second Language Learning and Teaching Conference*, J.M. Levis and K. LeVelle (eds.), 105–118, Ames, IA: Iowa State University.

Kang, O. 2012. Impact of rater characteristics and prosodic features of speaker accentedness on ratings of international teaching assistants' oral performance. *Language Assessment Quarterly* 9: 249–269.

Kelly, L.G. 1969. *25 Centuries of Language Teaching*, Rowley, MA: Newbury House.

Kelly, G. 2000. *How to Teach Pronunciation*, Harlow, England: Pearson Education.

Kennedy, S. and Trofimovich, P. 2008. Intelligibility, comprehensibility, and accentedness of L2 speech: the role of listener experience and semantic context. *Canadian Modern Language Review* 64: 459–489.

Kennedy, S. and Trofimovich, P. 2010. Language awareness and second language pronunciation: a classroom study. *Language Awareness* 19(3): 171–185. doi: 10.1080/09658416.2010.486439.

Kenworthy, J. 1987. *The Teaching of English Pronunciation*, New York: Longman.

Kenyon, J.S. 1928. Correct pronunciation. *American Speech* 4: 150–153.

Kingdon, R. 1958a. *The Groundwork of English Intonation*. London: Longmans.

Kingdon, R. 1958b. *The Groundwork of English Stress*. London: Longmans.

Ladefoged, P. 2006. *A Course in Phonetics*, 5th edition, Boston, MA: Thomson Wadsworth.

Lado, R., Fries, C.C., and Robinett, B.W. 1954. *English Pronunciation*. Ann Arbor, MI: The University of Michigan Press.

Lane, L. 2005. *Focus on Pronunciation* (books 1, 2, and 3), White Plains, NY: Pearson Longman.

Lane, L. 2010. *Tips for Teaching Pronunciation: A Practical Approach*, White Plains, NY: Pearson Longman.

Laroy, C. 1995. *Pronunciation*, New York: Oxford University Press.

Leather, J. 1983. State of the art: second-language pronunciation learning and teaching. *Language Teaching* 16: 198–219.

Leather, J. 1999. Second-language speech research: an introduction. *Language Learning* 49: 1–56.

Levis, J. 1999. Intonation in theory and practice, revisited. *TESOL Quarterly* 33: 37–63.

Levis, J.M. and Pickering, L. 2004. Teaching intonation in discourse using speech visualization technology. *System* 32: 505-–524.

Llurda, E. 2000. Effects of intelligibility and speaking rate on judgements of non-native speakers' personalities. *IRAL* 38: 289–299.

Long, M. and Crookes, G. 1991. Three approaches to task-based syllabus design. *TESOL Quarterly* 26: 27–55.

Lord, G. 2008. Podcasting communities and second language pronunciation. *Foreign Language Annals* 41(2): 364–379. doi: 10.1111/j.1944-9720.2008.tb03297.x.

Low, E.L. 2006. A cross-varietal comparison of deaccenting and given information: implications for international intelligibility and pronunciation teaching. *TESOL Quarterly* 40, 739–761.

Macdonald, S. 2002. Pronunciation views and practices of reluctant teachers. *Prospect* 17: 3–18.

Macdonald, D., Yule, G., and Powers, M. 1994. Attempts to improve English L2 pronunciation: the variable effects of different types of instruction. *Language Learning* 44: 75–100.

Marks, J. 2007. *English Pronunciation in Sse* (elementary), Cambridge: Cambridge University Press.

Marcel, C. 1853. *Language as a Means of Mental Culture and International Communication or Manual of the Teacher and the Learner of Languages*, London: Chapman and Hall.

Matsuura, H. 2007. Intelligibility and individual learner differences in the EIL context. *System* 35: 293–304.

McCutcheon, R.J. 1939. Correct pronunciations. *American Speech* 14: 159–160.

Miller, S.F. 2006. *Targeting Pronunciation: Communicating Clearly in English*, 2nd edition, New York: Houghton Mifflin.

Morley, J. 1984. *Listening and Language Learning in ESL*, Orlando, FL: Harcourt Brace Jovanovich.

Morley, J. (ed.). 1987. *Current Perspectives on Pronunciation: Practices Anchored in Theory*, Washington, DC: TESOL.

Morley, J. 1991. The pronunciation component in teaching English to speakers of other languages. *TESOL Quarterly* 25: 481–520.

Morley, J. 1994. A multidimensional curriculum design for speech-pronunciation instruction. In: *Pronunciation Pedagogy and Theory: New Views, New Directions*, J. Morley (ed.), 64–91, Alexandria, VA: TESOL.

Munro, M.J. and Derwing, T.M. 1995. Foreign accent, comprehensibility and intelligibility in the speech of second language learners. *Language Learning* 45: 73–97.

Munro, M.J. and Derwing, T.M. 1998. The effects of speaking rate on listener evaluations of native and foreign-accented speech. *Language Learning* 48: 159–182.

Munro, M.J. and Derwing, T.M. 2006. The functional load principle in ESL pronunciation instruction: an exploratory study. *System* 34: 520–-531.

Munro, M.J. and Derwing, T.M. 2011. Research timeline: the foundations of accent and intelligibility in pronunciation research. *Language Teaching* 44: 316–327.

Munro, M.J., Derwing, T.M., and Morton, S.L. 2006. The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition* 28: 111–131.

Murphy, J. 1997. Phonology courses offered by MATESOL programs in the US. *TESOL Quarterly* 31: 741–764.

Murphy, J. 2013. *Teaching Pronunciation*, Alexandria, VA: TESOL International Association.

Naiman, N. 1987. Teaching pronunciation communicatively. In: *The Teaching of Pronunciation: An Introduction for Teachers of English as a Second Language*, Special issue, P. Avery and S. Ehrlich (eds.). *TESL Talk* 17: 141–147.

Nunan, D. and Miller, L. 1995. *New Ways in Teaching Listening*, Alexandria, VA: TESOL.

O'Connor, J.D. and Arnold, G.F. 1961). *Intonation of Colloquial English*, London: Longmans.

Osburne, A.G. 2003. Pronunciation strategies of advanced ESOL learners. *International Review of Applied Linguistics* 41: 131–143.

Palmer, H.E. 1917. *The Scientific Study and Teaching of Languages*, London: Harrap.

Palmer, H.E. 1924. *The Grammar of Spoken English, on a Strickly Phonetic Basis*, Cambridge: Heffer.

Passy, P. 1888. Our revised alphabet. *The Phonetic Teacher*: 57–60.

Pennington, M. and Richards, J.C. 1986. Pronunciation revisited. *TESOL Quarterly* 20: 207–225.

Pica, T. 1984. Pronunciation activities with an accent on communication. *English Teaching Forum* 22: 2–6.

Pickering, L. 2001. The role of tone choice in improving ITA communication in the classroom. *TESOL Quarterly* 35: 233–255.

Pickering, L. 2004. The structure and function of intonational paragraphs in native and nonnative speaker instructional discourse. *English for Specific Purposes* 23: 19–43.

Pickering, L., Hu, G., and Baker, A.A. 2012. The pragmatic function of intonation: cueing agreement and disagreement in spoken English discourse and implications for ELT. In: *Pragmatics, Prosody and English Language Teaching*, J. Romeo (ed.), 199–218, New York: Springer-Verlag.

Pike, K.L. 1945. *The Intonation of American English*, Ann Arbor, MI: The University of Michigan Press.

Poedjosoedarmo, G. 2003. *Teaching Pronunciation: Why, What, When, and How*, Singapore: RELC Portfolio Series.

Pourhosein Gilakjani, A. 2012. The significance of pronunciation in English language teaching. *English Language Teaching* (Canadian Center of Science and Education) 5: 96–107.

Prator, C. 1951. *Manual of American English Pronunciation for Adult Foreign Students*, New York: Holt, Rinehart, & Winston.

Prator, C., and Robinett, B.W. 1985/1972/1957. *Manual of American English Pronunciation*, 4th edition (3rd edition, 1972; 2nd edition, 1957), New York: Holt, Rinehart & Winston.

Predergast, T. 1864. *The Mastery of Languages or The Art of Speaking Foreign Tongues Idiomatically*, London: R. Bently.

Reed, M. and Michaud, C. 2005. *Sound Concepts: An Integrated Pronunciation Course*, New York: McGraw-Hill.

Richards, J.C. and Rodgers, T.S. 2001. *Approaches and Methods in Language Teaching*, 2nd edition, Cambridge: Cambridge University Press.

Rogerson-Revell, P. 2011. *English Phonology and Pronunciation Teaching*, London: Continuum.

Saito, K. 2007. The influence of explicit phonetic instruction on pronunciation teaching in EFL settings: the case of English vowels and Japanese learners of English. *The Linguistics Journal* 3: 16–40.

Saito, K. 2011. Identifying problematic segmental features to acquire comprehensible pronunciation in EFL settings: the case of Japanese learners of English. *RELC Journal* 42: 363–378.

Saito, K. 2013. Re-examining effects of form-focused instruction on L2 pronunciation development: the role of explicit phonetic information. *Studies in Second Language Acquisition* 35(1): 1–29. doi: 10.1017/S0272263112000666.

Saito, K. and Lyster, R. 2012a. Effects of form-focused instruction and corrective feedback on L2 pronunciation development of /r/ by Japanese learners of English. *Language Learning* 62: 595–683.

Saito, K. and Lyster, R. 2012b. Investigating the pedagogical potential of recasts for L2 vowel acquisition. *TESOL Quarterly* 46: 387–398.

Sauveur, L. 1874. *Introduction to the Teaching of Living Languages without Grammar or Dictionary*, Boston: Schoenhof and Moeller.

Scales, J., Wennerstrom, A., Richard, D., and Wu, S.H. 2006. Language learners' perceptions of accent. *TESOL Quarterly* 40: 715–738.

Setter, J. 2006. Speech rhythm in world Englishes: the case of Hong Kong. *TESOL Quarterly* 40: 763–782.

Setter, J. and Jenkins, J. 2005. State-of-the-art review article: pronunciation. *Language Teaching* 38: 1–17.

Sifakis, N.C. and Sougari, A.-M. 2005. Pronunciation issues and EIL pedagogy in the periphery: a survey of Greek state school teachers' beliefs. *TESOL Quarterly* 39: 467–488.

Sledd, J. 1960. Review of the book *The Groundwork of English Intonation* by Roger Kingdon. *Language* 36(1): 173–179.

Sweet, H. 1880–1881. Sound notation. *Transactions of the Philological Society*: 177–235.

Tanner, M.W. and Landon, M.M. 2009. The effects of computer-assisted pronunciation readings on ESL learners' use of pausing, stress, intonation, and overall comprehensibility. *Language Learning and Technology* 13: 51–65.

Terrell, T. 1977. A natural approach to second language acquisition and learning. *Modern Language Journal* 6: 325–337.

Trofimovich, P. and Baker, W. 2006. Learning second language suprasegmentals: effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition* 28: 1–30.

Trofimovich, P. and Gatbonton, E. 2006. Repetition and focus on form in L2 Spanish word processing: implications

for pronunciation instruction. *The Modern Language Journal* 90: 519–535.
Trofimovich, P., Lightbown, P.H., Halter, R.H., and Song, H. 2009. Comprehension-based practice: the development of L2 pronunciation in a listening and reading program. *Studies in Second Language Acquisition* 31: 609–639.
Underhill, A. (1994) *Sound Foundations: Living Phonology*, Oxford: Heinemann.
Ur, P. 1988. *Grammar Practice Activities*, New York: Cambridge University Press.
Van Ek, J.A. 1973. *The '"Threshold Level" in a Unit/Credit System*, Strasbourg: Council of Europe.
Vietor, W. 1882. *Der Sprachunterricht muss umkehren! Ein beitrag zur Überbürdunsfrage*, Under the pseudonym Quousque Tandem, Heilbronn: Henninger.
Walker, R. 2010. *Teaching the Pronunciation of English as a Lingua Franca*, Oxford: Oxford University Press.
Wells, J.C. 1998. Obituary: "Doc" O'Connor: Sound scholar, *The Guardian*, 13 August, p. 018.
Wennerstrom, A. 1994. Intonational meaning in English discourse: a study of non-native speakers. *Applied Linguistics* 15: 399–421.
West, M. 1927/1935. *New Methods Readers*, Calcutta: Longmans, Green.
Widdowson, H.G. 1972. The teaching of English as communication. *English Language Teaching* 27: 15–19.
Wilson, G.P. 1937. Standards of correct pronunciation. *Quarterly Journal of Speech* 23: 568–576.
Wong, R. 1987. *Teaching Pronunciation: Focus on English Rhythm and Intonation*, Englewood Cliffs, NJ: Prentice-Hall Regents.
Zielinski, B.W. 2008. The listener: no longer the silent partner in reduced intelligibility. *System* 36: 69–84.

Part II Describing English Pronunciation

4 Segmentals

DAVID DETERDING

Introduction

The development of an alphabetic system of writing is one of the major milestones in the evolution of Western civilization, allowing a huge range of words to be shown using a small set of symbols. However, the 26 letters in the Roman alphabet are not sufficient to represent all the sounds of English in a straightforward manner, particularly as there are only five vowel letters while there are many more vowel sounds in all varieties of English. As a result, additional symbols have been developed to represent the segmental sounds accurately, not just for English but for all human languages, using the International Phonetic Alphabet (IPA).

However, it is unclear how many consonants and vowels there actually are in English and also how they should best be represented. Some of this uncertainty arises because of the existence of different accents, so that, for example, some people differentiate *which* from *witch*, so these speakers may have one more consonant than those for whom these two words are homophones, and the vowel in words such as *hot* and *calm* is different in many varieties of British English but the same for most speakers in the United States, which means that there is an extra monophthong vowel in British English. In addition, while use of IPA symbols for the consonants and vowels is certainly a convenient way of showing how words are pronounced, it is not clear whether these symbols in fact accurately reflect the true nature of English sounds, or whether some other kind of representation might be more appropriate, maybe using distinctive features such as [+voice] and [–nasal] or else by showing components of the sounds such as voicing and nasality on separate tiers. Discussion of the inventory of English segments, the symbols that are used to represent them, and also the nature of the phonological representation of consonants and vowels can provide valuable insights into the sound system of English.

In this chapter, after describing the emergence of a standard for the pronunciation of English, I will provide an overview of the symbols that are adopted to represent

The Handbook of English Pronunciation, First Edition. Edited by Marnie Reed and John M. Levis.

the vowels and consonants of English, including a comparison between the symbols that are typically used in Britain and in North America, and also the lists of sounds that are generally considered to constitute the inventory of phonemes in each variety. I will then briefly consider alternative nonsegmental models of pronunciation, such as the use of distinctive features and also autosegmental phonology, before discussing nonprescriptive ways of representing the segmental phonemes of English in order to derive a system that is not linked to any one standard that is promoted as the norm. There are many ways of pronouncing English, and some speakers around the world prefer no longer to be constrained by the symbols that are more appropriate for representing a standard accent that comes from Britain or North America, so it is valuable to consider how we can show the sounds of English without linking the representation to one accent.

The emergence of standard pronunciation

In the time of Shakespeare at the end of the sixteenth century, there was no established norm for the pronunciation of English, and it was only in the following centuries that a standard gradually emerged, largely based on the pronunciation of educated people in London and the south-east of England (Mugglestone 2003). Selection of one particular accent as the standard for pronunciation resulted in that accent having a privileged status while other styles of speaking were often disparaged, even though linguistically there is nothing inherently superior in one variety over another.

In 1755, one and a half centuries after the time of Shakespeare, when Dr. Samuel Johnson was compiling his dictionary, he still concluded that sounds were highly volatile and any attempt to fix them was futile; yet within a few decades, people such as John Walker and Thomas Sheridan were making substantial careers out of writing books and presenting well-attended lectures about elegant and correct pronunciation. Indeed, in his *Critical Pronouncing Dictionary* published in 1791, John Walker asserted that deviations from the elegant patterns of speech of genteel people were "ridiculous and embarrassing" (Mugglestone 2003: 23).

Of course, attempts to fix the pronunciation of English have proved somewhat elusive, just as Dr. Johnson predicted, and it is instructive to note that many features that are firmly established as standard in RP British English today, including the use of /ɑː/ in words such as *fast* and *bath* as well as the loss of postvocalic /r/ in words such as *morn* and *sort*, were condemned as "vulgar" or even "atrocities" by many people in the nineteenth century. Indeed, in a review written in 1818, the poet John Keats was condemned as uneducated and lacking in imagination partly because he rhymed *thoughts* with *sorts*, but rhyming these two words would nowadays be regarded as perfectly standard in British English (Mugglestone 2003: 78, 88).

In fact, two alternative standards for the pronunciation of English have emerged, one derived from the educated speech of the south-east of England and the other based on that of North America. These alternative standards give rise to

a number of issues about how many consonants and vowels there are in English and also how they should be represented, as I will outline in the following sections.

In the modern world, there are many valuable reference works showing the pronunciation of English words, especially the two principle pronouncing dictionaries, Wells (2008) and Jones et al. (2003). However, modern lexicographers usually see their role as descriptive rather than prescriptive, documenting a range of possible pronunciations for many words and sometimes offering substantial evidence for the patterns of pronunciation they report. Indeed, throughout his dictionary, John Wells provides data from a series of detailed surveys about pronunciation preferences. For example, *forehead* used to be pronounced with no /h/ in the middle, but Wells (2008: 317) reports that 65% of British respondents and 88% of Americans now prefer a pronunciation with a medial /h/. Furthermore, the percentage is highest among younger respondents, suggesting it is becoming established as the norm. We can say that the pronunciation of this word has changed because of the influence of its spelling (Algeo 2010: 46). Similarly, 27% of British respondents and 22% of Americans now state that they prefer *often* with a medial /t/, another trend that seems to be growing among younger people, though the fact that only a minority currently have a /t/ in this word suggests that this pronunciation is less advanced in becoming the norm (Wells 2008: 560). This work in conducting preference surveys to provide in-depth snapshots into changing patterns of speech represents a welcome effort to reflect pronunciation as it actually is rather than trying to impose some preconceived notion of what it should be. The fact that the pronunciation of words such as *forehead* and *often* seems to be shifting also illustrates that, even though standards nowadays exist for the pronunciation of English, the details are always undergoing change.

The International Phonetic Association (IPA)

The International Phonetic Association was established in 1886 with the aim of developing a set of symbols that could be used for representing all the sounds of the languages of the world (IPA 1999: 3). As far as possible, the letters from the Roman alphabet were adopted to represent their familiar sounds, so [b] is the IPA symbol for the voiced plosive produced at the lips and [s] is the symbol for the voiceless fricative produced by the tip of the tongue against the alveolar ridge. This is consistent with the way these letters are generally used in the writing systems of most European languages. Some of the extra symbols needed for other sounds were taken from the Greek alphabet, with, for example, [θ] representing a voiceless dental fricative, and other symbols were created by altering the shape of an existing letter, so, for instance, [ŋ] represents the nasal sound produced at the velum. Inevitably, with only five vowel letters in the Roman alphabet, additional symbols were needed to represent the full range of vowel sounds that occur in the languages of the world, so, for example, [ɒ] is the symbol that was created to represent an open back rounded vowel.

The IPA chart now shows 58 basic consonants, 10 nonpulmonic consonants (for clicks, implosives, and ejectives), 10 other consonant symbols such as [w] and its voiceless counterpart [ʍ], which both involve two places of articulation (labial and velar), and 28 vowels, as well as a range of symbols for tones, other suprasegmentals, and diacritics. The IPA symbols are periodically updated, such as at the Kiel Convention in 1989, to reflect enhanced knowledge about languages around the world. Nevertheless, few fundamental changes were made at the Kiel Convention (Esling 2010: 681), as the IPA is now well established and allows phoneticians to describe and compare a wide range of different languages quite effectively.

One issue that might be questioned concerning the IPA symbols is the use of [a] to represent a front vowel while [ɑ] is a back vowel. This seems to be the only case where a variant of a common Roman letter represents something different – it might be noted, for example, that selection between [g] and [ɡ] does not indicate a different sound – and the occurrence of both [a] and [ɑ] can give rise to confusion. Indeed, some writers use [a] not for a front vowel but to represent an unspecified open vowel, or sometimes even a back vowel. Because of this, Roca and Johnson (1999: 128) decided to take "the bold step of departing from IPA doctrine" in using [æ] instead of [a] to represent a fully open front unrounded vowel. However, for the representation of vowel quality in a range of languages, other writers do not seem to have followed their lead in this matter, apart from in the description of English for which the open front vowel in a word such as *man* is indeed represented as /æ/. I will now discuss the symbols used to show the sounds of English.

Phonemes and allophones

In the discussion of the IPA in the previous section, the symbols were enclosed in phonetic square brackets: [ ]. This is because the discussion was dealing with language-independent sounds such as [b] and [s] rather than the sounds of any one language. However, when considering the inventory of sounds in English, the consonant and vowel phonemes are shown in phonemic slashes: //. First, however, let us consider what is meant by a phoneme.

A phoneme is a contrastive sound in a language, which means that changing from one phoneme to another can create a new word (Laver 1994: 38). For example, the sound at the start of the word *pat* is represented as /p/, but if this /p/ is replaced with /b/, we get a different word, *bat*. We call *pat* and *bat* a minimal pair, and the existence of a minimal pair such as this confirms that /p/ and /b/ are different phonemes of English. Similarly, *save* and *safe* constitute a minimal pair, the existence of which demonstrates that /v/ and /f/ are different phonemes of English.

Another entity that should be introduced is the allophone. Allophones are variants of phonemes. For example, the /k/ at the start of *kit* is similar but not quite the same as the /k/ in *cat*, because the former is pronounced a little further forward in the mouth as a result of the influence of the following vowel (Ladefoged

and Johnson 2011: 77). We show allophones in phonetic square brackets and we use diacritics to indicate the fine details of the pronunciation, so the sound at the start of *kit* can be shown as [k̟] to indicate that it is produced further forward in the mouth than the [k̠] in *cat.* Allophones cannot create a new word because their occurrence can be predicted from where they are in a word and what occurs before them and after them (Gussenhoven and Jacobs 2011: 62).

I will now consider the inventory of phonemes in English, starting with consonants and then dealing with vowels.

Representing the consonants of English

Consonants can be described in terms of three basic parameters: whether they are voiced or voiceless; where in the vocal tract they are pronounced; and how they are pronounced. We can therefore say, for example, that the /p/ sound at the start of *pit* is a voiceless bilabial plosive. In other words, the vocal folds are not vibrating when it is produced (so it is voiceless); it is produced with both lips (it is bilabial); and it is articulated by means of a sudden release of the closure (it is a plosive).

It is generally agreed that there are 24 consonant phonemes in English, as shown in Table 4.1. The columns in Table 4.1 represent the place of articulation, so /p/ is presented in the column for bilabial sounds; the rows indicate the manner of articulation, so /p/ is in the row for plosives. Symbols on the left of any cell are voiceless, while those on the right of a cell are voiced, so /p/ is on the left of its cell to show it is voiceless, while its voiced equivalent, /b/, is on the right of the same cell. Many cells only have a single symbol. For example, /m/ appears on the right of the cell for bilabial nasal, but there is no voiceless equivalent as voiceless nasals do not occur in English.

One issue with the consonants as shown in Table 4.1 concerns /w/, which actually has two places of articulation, bilabial and velar, though it is only shown in the bilabial column. In fact, as mentioned above, in the IPA chart [w] is listed under "other symbols" rather than in the main table of consonants (IPA 1999: ix), because of this anomaly in its having dual articulation.

One might note that the use of /r/ to represent the postalveolar approximant is not quite accurate according to the IPA chart, in which [r] represents a trill, not an approximant. Strictly speaking, the postalveolar approximant should be shown as /ɹ/ rather than /r/. However, the more familiar symbol /r/ is adopted here, following the usual practice of scholars such as Cruttenden (2008: 157) and Roach (2009: 52).

I will discuss three issues regarding the inventory of 24 English consonants that are shown in Table 4.1: why /tʃ/ and /dʒ/ are considered as phonemes; whether /ŋ/ is really a phoneme in English; and whether /ʍ/, the voiceless counterpart of /w/, might be included.

The phonemes /tʃ/ and /dʒ/ consist of two consecutive sounds, a plosive followed by a fricative. So why do we classify them as single phonemes rather than two separate consonants? After all, *tax* /tæks/ is considered to have two

Table 4.1 The 24 consonant phonemes of English, classified according to place and manner of articulation.

		Place of articulation							
		Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Palatal	Velar	Glottal
Manner of articulation	Plosive	p b			t d			k g	
	Fricative		f v	θ ð	s z	ʃ ʒ			h
	Affricate					tʃ dʒ			
	Nasal	m			n			ŋ	
	Lateral				l				
	Approximant	w				r	j		

consonants at the end, /k/ followed by /s/, so why do we regard *catch* /kætʃ/ as having just one consonant at the end, /tʃ/, rather than /t/ followed by /ʃ/? Also, why is /tʃ/ shown in Table 4.1 while /ks/ is not? One factor here is that /tʃ/ and /dʒ/ are the only affricates that can occur at the start of a syllable, in words like *chip* /tʃɪp/ and *jug* /dʒʌg/, so in this respect they behave differently from other sequences of a plosive and fricative in English. For example, */ksɪp/ and */pfet/ are not well-formed words in English. (Here I am using the '*' symbol to indicate that a sequence of sounds is not well formed.) In addition, /tʃ/ and /dʒ/ are generally felt by users of English to be single consonants (Wells 1982: 48).

Now let us consider /ŋ/. Before /k/ and /g/, /ŋ/ occurs and we never find /n/. As mentioned above, if a sound can be predicted from the surrounding sounds, then it should be regarded as an allophone rather than a phoneme. Therefore it seems that if /ŋ/ might actually be regarded as an allophone of /n/, then it should be shown as [ŋ] (an allophone) rather than as the phoneme /ŋ/ (Roach 2009: 51). However, in some words, such as *sung*, /ŋ/ occurs without a following /k/ or /g/, and indeed there are minimal pairs such as *sung* /sʌŋ/ and *sun* /sʌn/ in which /ŋ/ contrasts with /n/. One possibility here is to suggest that *sung* actually has a /g/ after the nasal consonant, but this /g/ is silent as it is deleted when it occurs following a nasal consonant at the end of a word. However, suggesting the existence of silent underlying sounds is a level of abstraction that is generally avoided in representing the sounds of English, and this is why most writers prefer to regard /ŋ/ as a phoneme.

Finally, let us consider whether /ʍ/, the voiceless counterpart of /w/, should be included in Table 4.1. For some speakers, *which* and *witch* constitute a minimal pair: the first starts with /ʍ/ while the second starts with /w/. Therefore, should /ʍ/ be included in the inventory of English consonants? It is not included because only a minority of speakers nowadays have this sound. Wells (2008: 898) reports that only 23% of British speakers have /ʍ/ at the start of *white*, and for younger speakers the number is less than 10%, though the number is probably rather higher in North America (Wells 1982: 229).

Variation in the consonant symbols

Representation of the consonants of English using the IPA symbols listed in Table 4.1 is fairly standard, though there remain some differences between British and American usage. In particular, many writers in America (e.g., Fromkin and Rodman 1993; Finnegan 1994) use the 'hacek' symbols /š, ž, č, ǰ/ instead of the respective IPA symbols /ʃ, ʒ, tʃ, dʒ/. One advantage of using the hacek symbols is that /č/ and /ǰ/ clearly represent the affricates as single phonemes, which (as mentioned above) reflects the intuition of most speakers. In addition, some writers prefer the symbol /y/ instead of /j/ for the palatal approximant that occurs at the beginning of words such as *yes* and *yam*. Notice that the use of /y/ for the palatal approximant mirrors the English spelling, which is an advantage for people who are primarily interested in representing the sounds of English and are not too

concerned with the pronunciation of other languages. However, for cross-linguistic comparisons, it is best to use /j/ for the English approximant, as the IPA symbol [y] actually indicates a front rounded vowel such as that found in the French word *tu* ('you').

Representing the monophthong vowels of English

The quality of a vowel is usually described in terms of three basic variables: open/close; front/back; and rounded/unrounded. The first two depend on the position of the highest point of the tongue when producing the vowel. If the tongue is high in the mouth, we describe the vowel as close, while if it is low in the mouth, we say that the vowel is open; if the tongue is towards the front of the mouth, we describe the vowel as front, while if it is bunched at the back of the mouth, we say that it is a back vowel. The third variable depends on whether the lips are rounded or not. For example, the vowel in *food* (represented by the symbol /uː/ by most people in Britain, though many in North America prefer to show it as /u/) can be described as close back rounded, as the tongue is close to the roof and at the back of the mouth and the lips are rounded, while /æ/, the vowel in *man*, is open front unrounded, as the jaw is nearly fully open, the tongue is at the front of the mouth, and the lips are not rounded. Many scholars (e.g., Harrington 2010: 84) have suggested that these variables, particularly open/close and front/back, are in fact related more closely to the acoustics of the vowel rather than its articulation, as there is considerable variation in the ways that different speakers produce the same vowel. Nevertheless, the traditional labels provide an effective way of describing the quality of vowels even if they do not in fact reflect their actual articulation very closely.

The quality of the vowels can be shown on a vowel quadrilateral such as that in Figure 4.1, in which the front vowels are towards the left while the back vowels are on the right and close vowels are at the top while open vowels are near the bottom. This two-dimensional figure does not show rounding, but in English /uː/, /ʊ/, /ɔː/, and /ɒ/ are all rounded. The eleven monophthong vowels of British English that occur in stressed syllables are included in this figure. The position of the symbols, and also the shape of the vowel quadrilateral, are as shown in Roach (2009: 13 and 16).

One vowel of English that is omitted from Figure 4.1 is the schwa /ə/, because it can never occur in stressed syllables. If it were included, it would occupy the same position as /ɜː/, and this raises the issue of whether a separate symbol should be used for /ɜː/ and /ə/ or if the former should instead be shown as /əː/, i.e., as a long version of /ə/. The rationale for adopting a different symbol is that the other long/short vowel pairs, such as /iː/ and /ɪ/, are represented by means of distinct symbols as well as the length diacritic, so it would be an anomaly if /ɜː/ and /ə/ were an exception.

These symbols are fairly well established, though some people use /ɛ/ instead of /e/ for the vowel in a word such as *pet* because this vowel is usually nearly open-mid. Indeed, Schmitt (2007) makes a strong case that /ɛ/ is preferable.

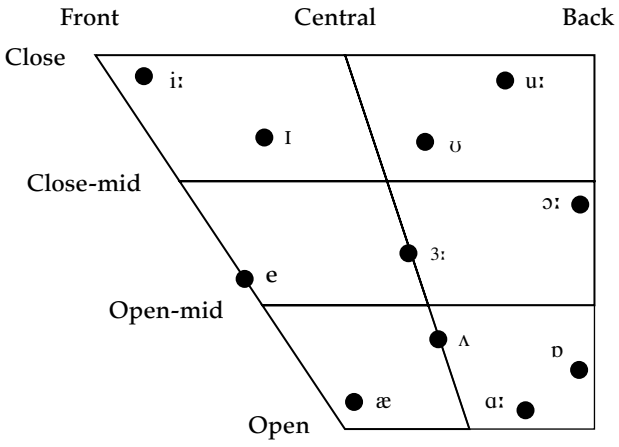


Figure 4.1 The monophthong vowels of British English.

One other issue regarding the use of symbols is whether the length diacritic should be used with /iː, ɑː, ɔː, uː, ɜː/. Some people omit this diacritic on the basis that these vowels are tense rather than long, and the tense vowels may actually be shorter in duration in many situations than the lax vowels, depending on the phonological environment and speaking rate. For example, the tense vowel /iː/ in *beat* may in fact be shorter than the lax vowel /ɪ/ in *bid*, because the final voiceless consonant in *beat* shortens the duration of the preceding vowel (Roach 2009: 28).

One might also note that /ɒ/ is absent from many varieties of American English (Wells 1982: 273), because the majority of people in the United States pronounce words such as *hot* and *shop* with /ɑː/ rather than /ɒ/ (though most speakers in Canada have /ɒ/ in these words). One other difference is that the mid central vowel in North America generally has r-coloring so it is sometimes shown as /ɝː/ (Wells 2008).

The location of some of the vowels in Figure 4.1 might be discussed further, in particular the exact positioning of /uː/. Acoustic measurements have suggested that /uː/ in modern RP Britain English is actually often more fronted than suggested by Figure 4.1 (Deterding 2006) and it seems that this is becoming increasingly true for younger speakers (Hawkins and Midgley 2005). However, like Roach (2009: 16), Wells (2008: xxiii) shows it as a back vowel and so does Cruttenden (2008: 127), who observes that a fronted variant mostly only occurs after the approximant /j/ in words such as *youth* and *cute*.

Diphthongs

The quality of monophthongs does not change very much during the course of the vowel. In contrast, diphthongs have a shifting quality. RP British English generally has eight diphthongs: five closing diphthongs /eɪ, aɪ, ɔɪ, əʊ, aʊ/, in which the

quality of the vowel moves from a relatively open vowel towards a more close one, and three centring diphthongs /ɪə, eə, ʊə/, in which the endpoint of the vowel is at the centre of the vowel quadrilateral.

The major differences for North American Englishes are that /əʊ/ is usually represented as /oʊ/ (suggesting a less front starting point) and, as the pronunciation of most speakers is rhotic, there are no centring diphthongs, because the vowels /ɪə, eə, ʊə/ in words such as *peer*, *pair*, and *poor* are a sequence of a monophthong followed by /r/ so the rhyme of these words is /ɪr/, /er/, and /ʊr/ respectively. In a few words, such as *idea*, which have /ɪə/ in RP British English, there is no potential final /r/, so in most North American Englishes this word has three syllables /aɪ diː ə/ while it just has two syllables /aɪ dɪə/ in British English (Wells 2008: 398).

One other issue in the inventory of diphthongs is that Ladefoged and Johnson (2011: 93) regard /ju/, the vowel in a word such as *cue*, as a phoneme of English. However, as they note, this makes it distinct from all the other diphthongs of English, as it is the only one in which the most prominent part is at the end, which is one reason why most people consider it as a sequence of the approximant /j/ and the monophthong /uː/ rather than a diphthong of English.

Two of the centring diphthongs in British English might be discussed further: /ʊə/ and /eə/. Many speakers in Britain nowadays have /ɔː/ rather than /ʊə/ in words such as *poor*, *sure*, and *tour*, so for 74% of people *poor* and *pour* are homophones (Wells 2008: 627). However, most speakers have /ʊə/ after /j/ in words such as *cure* and *pure*, so it seems that the /ʊə/ diphthong still exists for the majority of people in Britain.

For /eə/, many speakers have little diphthongal movement in this vowel, and Cruttenden (2008: 151) describes its realization as a long monophthong [ɛː] as "a completely acceptable alternative in General RP". One might therefore suggest that the vowel in a word such as *hair* could be represented as /ɛː/. Nevertheless, most writers continue to use the symbol /eə/ for this vowel because it is well established and we should be hesitant about abandoning a convention that is adopted in textbooks throughout the world whenever there are small shifts in actual pronunciation.

We might further ask whether there is actually a need to list any diphthongs in English, and indeed some writers prefer to show the vowel /aʊ/ in a word like *how* as /aw/ (i.e., a monophthong followed by an approximant). We might note that *say* is similar to *yes* spoken backwards and also that *my* is rather like *yum* said backwards, and if *yes* and *yum* are transcribed with an initial approximant, then it might seem to make sense similarly to represent *say* and *my* with a final approximant, as /sej/ and /maj/ respectively, though Wells (1982: 49) notes that it is uncertain if the vowel in *my* should be /maj/ or /mʌj/ or something else. Similarly, words such as *low* and *cow* might be shown as /low/ and /kaw/ respectively. If we show these words with /j/ and /w/ at the end, then there is no need to list closing diphthongs in the inventory of English vowels, as we only have monophthongs optionally followed by an approximant. However, this solution works better for a rhotic accent such as most varieties of North American English than a

non-rhotic accent such as RP British English, because RP has the additional centring diphthongs /ɪə, eə, ʊə/ in words such as *peer*, *pair*, and *poor*. So it seems that diphthongs are needed for representing RP, and if diphthongs are needed for the centring diphthongs, then we might as well show *say* and *my* with diphthongs as well.

We can further consider which sounds are classified as diphthongs. The vowels in words such as *day* and *go* are actually monophthongs in many varieties of English, including those of most speakers from Wales (Wells 1982: 382), Scotland (Wells 1982: 407), Singapore (Deterding 2007: 25) and many other places (Mesthrie and Bhatt 2008: 123–124). It is therefore not clear if it is appropriate that these two sounds should be classified as diphthongs just because RP speakers from Britain and many speakers in North America pronounce them that way. At the end of this paper, I will discuss nonprescriptive ways of referring to these vowels, using the keywords FACE and GOAT and thereby avoiding symbols such as /eɪ/ and /əʊ/, which make the assumption that they are diphthongs.

Feature-based representations of sounds

One issue with representing the sounds of English (or any other language) in terms of phonetic symbols is that it fails to reflect some regularities. For example, /p, t, k/ form a natural class of consonants, namely the voiceless plosives, while /m, s, j/ do not form a natural class, but this is not reflected by showing them as a list of symbols. It is not easy to write a rule to represent some phonological process unless there is some formal way of identifying natural classes of sounds. For instance, when the voiceless plosives occur at the start of a stressed syllable (e.g., *pan*, *tough*, *kill)*, they are usually aspirated, which means that a little puff of air occurs after they are released. However, when they occur after initial /s/ (e.g., *span*, *stuff*, *skill)*, they are not aspirated, and we cannot easily write a rule to show this using IPA symbols. Similarly, if we want to list the consonants that can occur after /k/ at the start of a syllable in English, we find only /r, w, j, l/ are permissible sounds in this position. However, this is not a random list of symbols, and it would be best to have a formal way of representing them.

One possible solution to this is to use distinctive features. For example, [+obstruent] represents a sound that is produced with a complete or partial blockage of the airflow, [+continuant] means that the blockage is not complete, [+delayed release] is used to represent the affricates, and [+voice] means that the sound is voiced, and we can then represent the voiceless plosives in terms of four distinctive features: [+obstruent –continuant –delayed release –voice]. Similarly, the approximants /r, w, j, l/ can be represented as [–obstruent +continuant] (Carr 1993: 65). Under this model, a phoneme such as /p/ does not really exist and is just the shorthand for a bundle of features. This was the approach proposed in the highly influential work *The Sound Pattern of English* (Chomsky and Halle 1968).

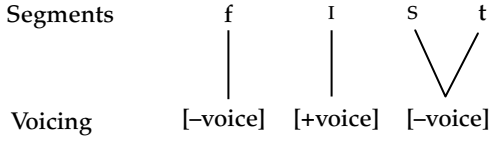
An essential goal of the work of Chomsky and Halle was to capture all the regularities that are found in English. However, this involved adopting highly abstract representations, such as a silent final /g/ in a word like *sung*

that I mentioned above. Moreover, some of the rules could become exceptionally complex. For example, Chomsky and Halle (1968: 52) proposed a rule that converts the /eɪ/ vowel in *sane* to the short vowel /æ/ in the first syllable of *sanity*, on the basis that this process occurs in a range of other words, including *vane/vanity* and *profane/profanity*, but in the attempts to capture this regularity, the representation of words ended up being substantially different from their surface realization. For this reason, the full rule-based framework proposed by Chomsky and Halle is not widely adopted by phonologists today in representing the phonology of English.

However, distinctive features are still often used to represent classes of sounds and to describe some of the phonological processes they undergo in speech. One issue that concerns these features is whether they are all binary, as with [±voice], or whether some of them might be unary, such as [labial] (Gussenhoven and Jacobs 2011: 74), but the details of this issue are beyond the scope of this brief overview of the segments of English.

Autosegmental representations

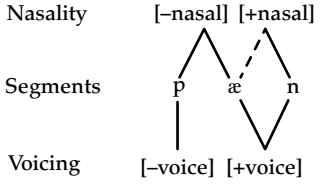
The use of distinctive features discussed in the previous section assumes that segment-sized phonemes may not be the fundamental phonological units of speech, and there is something smaller, namely the distinctive feature. One could alternatively propose that the segment is actually too small a unit for representing many aspects of phonology, and we should make use of features that extend over more than one segment. For example, in English, we do not find voiced consonants following voiceless ones at the end of a syllable, so /fɪst/ is fine but */fɪsd/ is not, as it involves voiced /d/ following voiceless /s/ in the coda of the syllable. In situations like this, it is redundant to show the voicing of both /s/ and /t/ independently, and maybe the [–voice] feature should be represented as extending over two successive segments. If voicing is separated from the rest of the segments and then shown in its own tier, we get something like this:



This representation accurately reflects the fact that the voicing feature can only change twice in an English syllable, from [–voice] to [+voice] and back to [–voice], so even in a syllable with seven segments such as *strengths* [streŋθs], the representation of voicing is still [–voice] [+voice] [–voice].

This kind of proposal, with separate tiers for different components of the pronunciation, was suggested by Goldsmith (1976) (though his work was mostly concerned with the representation of tones), and is termed autosegmental phonology.

Another feature that might be considered to belong on its own tier is nasality, and we might note that nasality does not always coincide with segment boundaries. For example, in a word such as *pan*, the vowel before a final nasal consonant becomes nasalized, but in fact only the end of the vowel gets nasalized. If nasality is represented in its own separate tier, as below, we can use a dotted line to show that the nasality of the final consonant extends over the previous sound while it does not prevent the first part of the vowel continuing to be non-nasal:



This representation of the word *pan* accurately reflects the fact that, for this word, both nasality and voicing only change once, even though there are three segments in it.

Nonprescriptive representations

Traditionally, a language such as English has been regarded as belonging to its native speakers, and IPA symbols for the standard pronunciation that native speakers use are assumed to be appropriate for representing the segments of the language. However, in the globalized modern world this assumption that native speakers own the English language has become problematic for many reasons. Firstly, it is hard to be sure what we mean exactly by a native speaker of English (McKay 2002: 28). If someone grows up speaking two languages equally well, are they a native speaker of both? And if someone only starts to speak English from the age of five but then develops perfect competency, are they a native speaker? Secondly, when English as a lingua franca (ELF) has become so widely used in the world and there are now far more non-native than native speakers using the language on a daily basis (Crystal 2003), can we continue to assume that ownership resides solely with its native speakers from places such as Britain and the United States?

In the past, some writers have suggested that native speakers are irrelevant for the description of ELF (Jenkins 2000). Others argue that native speakers may have a role in ELF corpora and thereby contribute to the analysis of patterns of usage that are discovered from those corpora (Seidlhofer 2011); indeed, more recently, when discussing the composition of ELF corpora, Jennifer Jenkins has acknowledged that native speakers do not need to be excluded from such corpora when they are talking to non-native speakers (Jenkins, Cogo, and Dewey 2011: 283). One way or another, whatever the status of native speakers in the description of ELF, there is nowadays a widely held view that non-native speakers should also have a

prominent voice in the evolution of standards for worldwide English, particularly proficient users in what Kachru (2005) has termed the outer-circle countries such as India, Nigeria, and Singapore, which were once colonies and where English continues to function widely as an official language.

This raises a question. How should we talk about the sounds of English without assuming that one style of pronunciation is "correct" or "better" than another? If proficient speakers of English around the world pronounce the sound at the start of a word such as *think* as [θ], [t], [s], or [f], how do we refer to this sound without assuming that one of these realizations (such as the dental fricative /θ/) is somehow better than the others? And if the vowel in a word such as *say* is a diphthong in some varieties of English but a monophthong in others, does it make any sense to represent it using the symbol /eɪ/ or, indeed, to list it as a diphthong as was done above in presenting the inventory of vowels of English?

The solution proposed by Wells (1982) is to use upper-case letters for many of the consonants and keywords written in small caps for the vowels. Using this system, we can talk about how the voiceless TH sound is realized in different accents, we can refer to processes such as T-glottaling and L-vocalization that affect consonants, and we can consider how vowels such as FACE and GOAT are pronounced around the world. Indeed, Wells introduced a set of 24 keywords for representing the vowels of English, and this system allows us to talk about differences between varieties of English in a nonprescriptive way. For example, we can say that TRAP is usually pronounced as [æ] and PALM is generally [ɑː], but the vowel in words such as *staff*, *brass*, *ask*, and *dance* that belong in the BATH lexical set may be pronounced as [ɑː] in the UK or as [æ] in the USA (Wells 1982: xviii). Note that this way of representing the pronunciation avoids giving a privileged status to either of the two accents.

This system is now quite widely adopted, though there are still some problems. For example, it would usually be assumed that the vowel in *bed* is DRESS and so in most varieties of English it is pronounced as [e] (usually written as [ɛ] for American English). However, in Singapore English the word *bed* actually rhymes with *made* and not with *fed* (Deterding 2005), which suggests that it may belong with FACE rather than DRESS. To some extent, therefore, we need to extend or modify the keywords. Deterding (2007: 12) introduced the keyword POOR to represent the vowel in words such as *poor*, *tour*, and *sure*, which in Singapore English are all pronounced as [ʊə]. The problem here is that the keyword for /ʊə/ is CURE, but in Singapore the word *cure* is usually pronounced as [kjɔː], and it seems unfortunate if the word *cure* does not have the CURE vowel. In fact, it is likely that further extensions and adaptations to the keywords may be needed to offer a comprehensive description of Englishes around the world.

Conclusion

Over the past two centuries, a standard pronunciation of English has emerged, originally based on the accent of educated people in London but later with an alternative standard based on the pronunciation of people in North America.

At the same time, the IPA symbols have been developed as a means of accurately representing all the sounds of human languages, and following from this, a fairly well-established set of symbols has emerged to represent the segmental sounds of English, even though there remain some differences between a few of the symbols that are used, particularly because of differences in the standard pronunciations of Britain and the United States. The adoption of an established set of symbols for indicating pronunciation is useful because there are substantial advantages in maintaining agreed conventions for the range of textbooks and reference materials that are produced today.

The use of the IPA segmental symbols may not accurately reflect some aspects of the structure and some of the processes that characterize English syllables, such as alternations in voicing in English syllables and the predictive assimilation of nasality for a vowel before a nasal consonant. However, there seems little chance that alternative representations, such as those based on distinctive features or tier-based autosegmental phonology, will displace the convenient, widely understood, and highly flexible IPA symbols to represent the sounds of English.

Perhaps the greatest challenge to the use of these well-established IPA symbols is the burgeoning spread of ELF and the corresponding need for nonprescriptive ways of referring to the sounds. Only time will tell how extensively writers will adopt the upper-case letters for consonants and small-caps keywords for vowels suggested by Wells (1982), whether the problems that remain in using these symbols will be ironed out, or if some alternative representation of the consonants and vowels of English will eventually emerge.

REFERENCES

Algeo, J. 2010. *The Origins and Development of the English Language*, 6th edition, Boston, MA: Wadsworth Cengage Learning.

Carr, P. 1993. *Phonology*, London: Macmillan.

Chomsky, N. and Halle, M. 1968. *The Sound Pattern of English*, Cambridge, MA: MIT Press.

Cruttenden, A. 2008. *Gimson's Pronunciation of English*, 7th edition, London: Hodder Education.

Crystal, D. 2003. *English as a Global Language*, Cambridge: Cambridge University Press.

Deterding, D. 2005. Emergent patterns in the vowels of Singapore English. *English World-Wide* 26: 179–197.

Deterding, D. 2006. The North Wind versus a Wolf: short texts for the description and measurement of English pronunciation. *Journal of the International Phonetic Association* 36: 187–196.

Deterding, D. 2007. *Singapore English*, Edinburgh: Edinburgh University Press.

Esling, J. 2010. Phonetic notation. In: *The Handbook of Phonetic Sciences*, 2nd edition, W.J. Hardcastle, J. Laver, and F. Gibbon (eds.), 678–702, Malden, MA: Wiley-Blackwell.

Finnegan, E. 1994. *Language: Its Structure and Use*, 2nd edition, Fort Worth, TX: Harcourt Brace.

Fromkin, V. and Rodman, R. 1993. *An Introduction to Language*, 5th edition, Fort Worth, TX: Harcourt Brace Jovanovich.

Goldsmith, J.A. 1976. An overview of autosegmental phonology. *Linguistic Analysis* 2(1): 23–68. Reprinted in *Phonological Theory: The Essential Readings*, J.A. Goldsmith (ed.), 137–161, Malden, MA: Blackwell.

Gussenhoven, C. and Jacobs, H. 2011. *Understanding Phonology*, 3rd edition, London: Hodder Education.

Harrington, J. 2010. Acoustic phonetics. In: *The Handbook of Phonetic Sciences*, 2nd edition, W.J. Hardcastle, J. Laver, and F. Gibbon (eds.), 81–129, Malden, MA: Wiley-Blackwell.

Hawkins, S. and Midgley, J. 2005. Formant frequencies of RP monophthongs in four age groups of speakers. *Journal of the International Phonetic Association* 35: 183–200.

IPA. 1999. *Handbook of the International Phonetic Association*, Cambridge: Cambridge University Press.

Jenkins, J. 2000. *The Phonology of English as an International Language*, Oxford: Oxford University Press.

Jenkins, J., Cogo, A., and Dewey, M. 2011. Review of developments in research into English as a lingua franca. *Language Teaching* 44(3): 281–315.

Jones, D., Roach, P., Hartman, J., and Setter. J. 2003. *Cambridge English Pronouncing Dictionary*, 16th edition, Cambridge: Cambridge University Press.

Kachru, B.B. 2005. *Asian Englishes: Beyond the Canon*, Hong Kong: Hong Kong University Press.

Ladefoged, P. and Johnson, K. 2011. *A Course in Phonetics*, 6th edition, Boston, MA: Wadsworth/Cengage Learning.

Laver, J. 1994. *Principles of Phonetics*, Cambridge: Cambridge University Press.

McKay, S.L. 2002. *Teaching English as an International Language*, Oxford: Oxford University Press.

Mesthrie, R. and Bhatt, R.M. 2008. *World Englishes: The Study of New Linguistic Varieties*, Cambridge: Cambridge University Press.

Mugglestone, L. 2003. *Talking Proper: The Rise of Accent as a Social Symbol*, 2nd edition, Oxford: Oxford University Press.

Roach, P. 2009. *English Phonetics and Phonology: A Practical Course*, 4th edition, Cambridge: Cambridge University Press.

Roca, I. and Johnson, W. 1999. *A Course in Phonology*, Oxford: Blackwell.

Schmitt, H. 2007. The case for the epsilon symbol (ɛ) in RP DRESS. *Journal of the International Phonetic Association* 37: 321–328.

Seidlhofer, B. 2011. *Understanding English as a Lingua Franca*, Oxford: Oxford University Press.

Wells, J.C. 1982. *Accents of English*, Cambridge: Cambridge University Press.

Wells, J.C. 2008. *Longman Pronunciation Dictionary*, 3rd edition, Harlow: Pearson Longman.

5 Syllable Structure

ADAM BROWN

Introduction

The topic of this chapter is one that is often overlooked in the description of language: syllables and their internal structure. The paper starts with a discussion of why the syllable is an important unit. The structure of the syllable is then examined, and English syllable structure is shown to be more complex than that of most other languages. After this preliminary basic explanation, various problems with it are investigated.

It is not possible in a paper of this length to go into all the rules that could be stated about English syllable structure. Instead, eight such rules are presented, as an indication of how complex English syllable structure is. While the word *rule* is used, these are generalizations about what does and does not occur, and they have fuzzy edges, rather than the stricter sense, as in the rules of football. The notion of whether syllables are regular (i.e., follow the rules) is distinguished from whether they are occurring as words or in words of English. We examine the way in which loanwords that are borrowed from one language to another are usually changed, if necessary, in order to conform to the syllable structure rules of the borrowing language.

Finally, the relevance of syllable structure to language teaching is explained.

Importance as a unit

Many people instinctively believe that the word is the most important unit in a language. One reason for this may be that they are influenced by spelling. Words are clearly units in spelling, as they have spaces or punctuation either side. In pronunciation, there are units that are larger and smaller than the word, and the syllable is one of the most important. In view of this, it is surprising that many of the descriptions of individual languages on Wikipedia and elsewhere analyze the

The Handbook of English Pronunciation, First Edition. Edited by Marnie Reed and John M. Levis.

vowel and consonant segments, and the suprasegmentals (stress, intonation, etc.), but say nothing about the syllable structure of the language.

There are several reasons why the syllable is an important unit. Some reasons relate to the fact that syllables are psychologically real to language users.

Syllabic writing systems (syllabaries)

There are some languages whose writing systems are based on the syllable, rather than the individual vowel and consonant sounds. The kana (hiragana and katakana) system of Japanese is the most familiar example of this (Bowring and Laurie 2004), but there are many others: Akkadian (Mesopotamia, extinct), Bopomofo (China, Taiwan), Cherokee (Southeast USA), Linear B (Greece, extinct), Mayan (Central America), Pahawh Hmong (Laos, Vietnam), and Vai (Liberia).

Ability to identify syllables

Everyone, regardless of their native language and its writing system, seems to be able to identify by and large how many syllables words contain (but see the section *Problems in syllabification* below). That is, it is a unit that people are consciously aware of. "[I]ndeed, explicit awareness of syllables [by children] has been shown to developmentally precede explicit awareness of phonemes" (Gnanadesikan 2008).

Importance to literacy

Literacy experts are agreed that an awareness of the syllables in a word, the sounds that make up the syllables, and of phenomena such as alliteration and rhyme (see below) are essential for efficient spellers of English (Carson, Gillon, and Boustead 2013; HearBuilder n.d.; Justice et al. 2013; Moats 2010; Moats and Tolman 2009; Wilson 2013).

Other reasons relate to the place of the syllable in linguistic analysis.

Hierarchy of phonological units

The syllable fits nicely into a hierarchy of phonological units. Features (such as [± voice], [± labial]) are present in segments (vowels and consonants). Segments make up syllables. Syllables combine into feet, units used in the analysis of speech rhythm, and tone groups, units used in intonational analysis, may be composed of one or more feet.

Stress

Stress in words is placed on syllables rather than individual vowel and consonant phonemes. For example, the noun *insight* and the verb *incite* have identical phonemes. The difference in their pronunciation is the stress placement,

on the first or second syllable: *insight* /ˈɪnsaɪt/, *incite* /ɪnˈsaɪt/, where /ˈ/ marks the start of the stressed syllable. Similarly, the intonational focus (tonic; see Chapters 8 and 10 on intonation) of an utterance falls on a particular syllable, rather than a phoneme or word. For instance, an utterance "It's absolutely ridiculous!" is likely to have the tonic (probably a fall from high to low) on the second syllable of *ridiculous*.

Combinations of phonemes

The syllable is the largest unit that is required for accounting for the combinations of phonemes in a language. For instance, is the sequence /mftfr/ possible in English? The answer is "yes", as in the phrase *triumphed frequently*. However, because of syllable constraints, there must be a syllable division between /mft/ and /fr/.

Allophonic realization rules

Many of the rules accounting for the occurrence of variants (allophones) of sounds (phonemes) can only be stated in terms of the syllable. For example, many accents of English distinguish between a "clear" /l/, with the tongue bunched upwards and forwards towards the hard palate, and a "dark" /l/, with the tongue bunched upwards and backwards towards the soft palate; in these accents, clear /l/ occurs at the beginning of syllables, as in *lick*, while dark /l/ occurs at the end of syllables, as in *kill*.

Differences between languages

Some differences between languages in the occurrence of sounds can only be stated in terms of the syllable and its structure. For example, the sounds /h/ and /ŋ/ occur in English and many other languages. However, in English /h/ can only occur at the beginning of a syllable, as in *help, behave* (/hɛlp, bɪ.heɪv/, where the dot marks the syllable division). However, there are languages, such as Arabic, Malay, and Urdu, where /h/ can occur at the ends of syllables, e.g., Malay *basah* /basah/ "wet". Notice that, in analysing syllable structure, we are talking about sounds (phonemes); the spelling is irrelevant. Thus, while many English words end in an *h* letter, this letter never represents an /h/ sound. It may be silent as in *messiah*, *cheetah*, or may work in combination with another letter, as in *th* (*path*), *ph* (*graph*), *sh* (*fish*), *gh* (*laugh*) and *ch* (*rich*). Note also that, while one-syllable (monosyllabic) words may often be given here as examples, they are given as examples of syllables, not of words. Also, differences between British English (BrE) and American English (AmE) are discussed where relevant.

Likewise, while /ŋ/ can occur in syllable-final position in English, as in *sang*, *ring*, it cannot occur in syllable-initial position. However, it can in other languages

(Anderson n.d. a, n.d. b) including Fijian, Malay/Indonesian, Māori, Thai, and Vietnamese; for instance, the Thai word for "snake" is /ŋu:/.

Some differences in sound combinations between languages can only be stated in terms of the syllable and its structure. For example, both English and German have the /p, f, l/ sounds. While English has the sequence /pfl/ in an example such as *hipflask*, this is only possible because there is a syllable (and morpheme) break between the syllable-final /p/ and the syllable-initial /fl/. In German, on the other hand, words can start with /pfl/ as in *pflegen* "to be accustomed (to doing something)". Therefore, while both languages have all three sounds, and both languages have sequences of /pfl/, in German these can all be in syllable-initial position, but in English they can only be across a syllable boundary in the middle of a word. This explains why English speakers find German words like this non-English and awkward to pronounce.

Structure of the syllable

The syllables that make up words are analyzed in terms of three positions. The minimal type of syllable is composed of only a vowel, e.g., *eye* /aɪ/, *owe* /oʊ/. The vowel is therefore considered a central part of any syllable, and is in peak position (also called *syllabic*, *syllable-medial*, and *nuclear*). Before the vowel, there may be one or more consonants, e.g., *tie* /taɪ/, *sty* /staɪ/. This position is known as the onset (also called *syllable-initial* or *releasing*). After the vowel, there may also be one or more consonants, e.g., *isle* /aɪl/, *isles* /aɪlz/. This position is known as the coda (also called *syllable-final*, *offset* or *arresting*). Table 5.1 shows various possibilities, where C stands for any consonant, V for any vowel, and O for an empty position. Syllables with an empty coda position are called *open* syllables, while *closed* syllables have final consonants.

Table 5.1 Syllable structure of various English words.

Word	*Onset*	*Peak*	*Coda*	*Formula*
eye		aɪ		OVO
isle		aɪ	l	OVC
tie	t	aɪ		CVO
tile	t	aɪ	l	CVC
isles		aɪ	lz	OVCC
sty	st	aɪ		CCVO
style	st	aɪ	l	CCVC
tiles	t	aɪ	lz	CVCC
styles	st	aɪ	lz	CCVCC

Complexity of English syllable structure

More than one consonant in either the onset or coda position is known as a cluster. Therefore the last five possibilities above contain clusters. The largest cluster in the onset position in English has three consonants, as in *string* /strɪŋ/. The largest cluster in the coda position has four consonants, as in *sculpts* /skʌlpts/. Often, large final clusters in English are simplified; for example, the /t/ of *sculpts* may be omitted (elided; see Chapter 9 on connected speech processes). Nevertheless, it is certainly there in an underlying sense.

We can thus represent English syllable structure by the formula $C_{0\text{-}3}\ V\ C_{0\text{-}4}$. Syllable structure formulae for other languages are given in Table 5.2.

The syllable structure of English is thus more complex than that of most languages. In an analysis of syllable structure, Maddieson (n.d.) divided languages of the world into three categories:

1. Simple syllable structure: those like Māori, with the formula $C_{0\text{-}1}\ V\ O$; that is, there are only two syllable types (OVO and CVO) and no clusters or final consonants.
2. Moderately complex syllable structure: those that can add one more consonant in either the initial or final position. This gives the formula $C_{0\text{-}2}\ V\ C_{0\text{-}1}$ and includes two-consonant initial clusters.
3. Complex syllable structure: those having more complex onsets and/or codas.

Of the 486 languages investigated, the distribution was:

Simple syllable structure	61
Moderately complex syllable structure	274
Complex syllable structure	151

English is clearly at the complex end of the syllable structure spectrum. For this reason, it is not surprising that English pronunciation is often simplified by foreign learners. Since learners are statistically likely to come from native languages with less complex syllable structures than English, they may find the clusters of English difficult and simplify them in various ways (see below). Similarly, in

Table 5.2 Syllable structure of various languages.

Māori	$C_{0\text{-}1}\ V\ O$	(i.e., only OVO and CVO syllables)
Cantonese	$C_{0\text{-}1}\ V\ C_{0\text{-}1}$	(i.e., no clusters)
Spanish	$C_{0\text{-}2}\ V\ C_{0\text{-}1}$	(i.e., initial clusters but no final clusters)
Arabic	$C_{0\text{-}1}\ V\ C_{0\text{-}2}$	(i.e., final clusters but no initial clusters)
Russian	$C_{0\text{-}4}\ V\ C_{0\text{-}4}$	(i.e., initial clusters and final clusters, both with up to 4 consonants)

the developmental speech of native children, consonants are first learnt individually before being combined into clusters; as a result clusters are simplified in the meantime before they are mastered (Williamson 2010).

Rhyme

There is a close bond between the peak and the coda, known together as the rhyme (sometimes spelt *rime*). Rhyme is an everyday concept in poetry, song lyrics, etc. Two syllables rhyme if they have identical peaks and codas. Imperfect rhyme means that the peaks and codas are not quite identical.The following limerick is said to have been written as a parody of Einstein's theory of relativity:

A rocket inventor named Wright /raɪt/
Once travelled much faster than light. /laɪt/
He departed one day /deɪ/
In a relative way /weɪ/
And returned on the previous night. /naɪt/

Wright, *light*, and *night* rhyme because they all end in /aɪt/, and *day* and *way* rhyme with /eɪ/. Multisyllable words rhyme if everything is identical from the vowel of the stressed syllable onwards, e.g., *computer* and *tutor* rhyme because they have identical /u:tə(r)/ from the stressed /u:/ vowel (/kəmˈpju:tə(r), ˈtju:tə(r)/, where /ˈ/ marks the start of the stressed syllable and (r) indicates that the /r/ is pronounced by some speakers and not by others (see *Rhoticity* below). Notice again that these phenomena relate to sounds; spelling is irrelevant to the discussion. Both these points are illustrated by the following limerick:

There was a young hunter named Shepherd /ˈʃɛpə(r)d/
Who was eaten for lunch by a leopard. /ˈlɛpə(r)d/
Said the leopard, replete, /rɪˈpli:t/
"He'd have gone down a treat /ˈtri:t/
If he had been salted and peppered!" /ˈpɛpə(r)d/

Onset

While the peak and coda are known as the rhyme, this leaves the onset as an independent element, and it has its own feature, known as alliteration. Syllables are said to alliterate if they contain identical onsets. Imperfect alliteration involves syllables whose onsets are not quite identical. Alliteration is a common feature of:

- Poetry and rhymes: *Round and round the rugged rock the ragged rascal ran*.
- Similes: *As busy as a bee; as dead as a doornail/dodo*.
- Idiomatic expressions: *Make a mountain out of a molehill; He who laughs last, laughs longest*.

- Names of commercial brands: *Dunkin' Donuts, PayPal, Bed Bath & Beyond*.
- Memorable names, both real and invented: *William Wordsworth, Charlie Chaplin, Donald Duck*.

Spoonerisms occur when the onsets of (the first syllables of) two words are transposed. Here are some slips attributed to Reverend Dr. William Archibald Spooner (1844–1930). The letters corresponding to the transposed sounds are underlined.

- *The Lord is a sh*oving *l*eopard.
- *I saw a student f*ighting *a l*iar *in the quadrangle.*
- *You have h*issed *all my m*ystery *lectures. You have t*asted *two w*orms. *Pack up your r*ags *and b*ugs, *and leave immediately by the t*own *d*rain. (The down train is the train to London.)

Spoonerisms are a type of slip of the tongue (Cutler 1982; Fromkin 1980). Slips of the tongue (or "tips of the slung") again show the division between the onset and the rhyme.

Problems in syllabification

The preceding discussion will hopefully have convinced you of the importance of syllables, but may also have led you to assume that the syllable is a simple unproblematic concept. This section examines some of the problems associated with syllables.

Number of syllables

Speakers may differ in their opinions as to the number of syllables particular words have. These differences may arise from various factors:

- ***Elision of /ə/:*** The schwa vowel /ə/ may be lost in certain environments, e.g., *comfortable* /kʌmfətəbəl/ (four syllables) or /kʌmftəbəl/ (three).
- ***Morphology:*** The word *evening* is one unit of meaning (morpheme) in *good evening* (usually /i:vnɪŋ/, two syllables), but is two morphemes (*even* + *ing*) in *evening out numbers* (usually /i:vənɪŋ/, three syllables).
- ***Spelling:*** The letter *a* in the spelling may lead speakers to believe there are three syllables in *pedalling* (thus /pɛdəlɪŋ/), but only two in *peddling* (thus /pɛdlɪŋ/).
- ***Long vowels + dark /l/:*** The vocalic nature of the darkness of dark /l/ may lead to different opinions about words such as *boil*.
- ***Triphthongs:*** Differences exist as to the syllabification of triphthongs, such as the /aɪə/ of BrE *fire* /faɪə/, as constituting one or two syllables.
- ***Compression:*** Sequences involving sounds such as /iə/ may be analysed as one or two syllables. For example, a word like *lenient* /li:niənt/ may be considered

in its fullest form to be three syllables; however, the /iə/ is often compressed (or smoothed; Wells 2008: 173–174) into one syllable, and may be reanalysed as /jə/, thus /li:njənt/.

Definition of the syllable

While the syllable may seem a clear entity, there is no universally agreed definition of the syllable. Three attempts to define the syllable (one articulatory, one acoustic, and one auditory) will be discussed here.

The prominence theory of the syllable is based on auditory judgments. Syllables correspond to peaks in prominence, usually corresponding to the number of vowels.

In the sonority theory (see, for example, Ladefoged and Johnson 2010), which is probably the most reliable and useful of the three attempts to define the syllable, syllables correspond to peaks in sonority. Sonority is the relative loudness (acoustic amplitude) of sounds compared with other sounds. This can be plotted on a scale of sonority (most sonorous first):

1. Low/open vowels, such as /æ, ɑ:/
2. Mid vowels, such as /ɛ, ɔ:/
3. High/close vowels, such as /i:, u:/ and semi-vowels /j, w/ (see below)
4. The lateral-approximant /l/
5. Nasal-stops, such as /m/
6. Voiced fricatives, such as /v/
7. Voiceless fricatives, such as /f/
8. Voiced oral-stops, such as /d/
9. Voiceless oral-stops and affricates, such as /t, tʃ/

While this works in most cases, there are exceptions. For instance, the word *believe* /bɪli:v/ has two vowels and two peaks of sonority. However, the word *spy* /spaɪ/ has one vowel, but the /s/ has greater sonority than the /p/; it also therefore has two peaks of sonority. Thus, whereas both words have two peaks of sonority, the first is clearly two syllables but the second only one. Many languages do not have initial clusters like /sp/ and they are often pronounced as two syllables by foreign learners. Similarly, instances involving syllabic consonants (see below) are counterexamples, e.g., *hid names* and *hidden aims* may involve the same sequence of phonemes /hɪdneɪmz/, but the first is two syllables while the second is three (involving a syllabic consonant; see below).

The articulatory chest pulse theory relates to the contraction of the intercostal muscles surrounding the lungs as they push air out during speech. It has been claimed (Gimson 1980: 56) that the number of chest pulses determines the number of syllables. This theory has been used most notably by Abercrombie (see, for example, 1967) in differentiating between syllable pulses and stress pulses, in order to formulate a theory of rhythm in speech.

Syllable boundaries

While speakers can usually tell how many syllables a word has, there may be confusion as to where one syllable ends and the next begins. For instance, in sequences where two vowels are separated by one or more consonants (e.g., *photon, pastor, outgrow, obstruct*), do these consonants belong with the first or the second syllable, or are they divided between them? Various writers (e.g., Wells 1990; Eddington, Treiman, and Elzinga 2013a, 2013b; Redford and Randall 2005) have researched this, investigating the features that correlate with syllabification preferences and proposing principles to account for them. Reasonably uncontroversial principles are the following:

- Syllable boundaries cannot divide the affricates /tʃ, dʒ/.
- Syllable divisions cannot create clusters that are otherwise impermissible. Thus, *panda* can be considered /pæn.də/ or / pænd.ə /, but not / pæ.ndə /, as /nd/ is an impermissible initial cluster.
- Syllable boundaries occur at morpheme boundaries. For instance, *loose-tongued*, regardless of whether you think it should be written as one word, hyphenated, or two words, could be analyzed as /lu:.stʌŋd/ or /lu:st.ʌŋd/ without breaking cluster constraints; however, analysts would always break it at the morpheme division: /lu:s.tʌŋd/.

However, that still leaves a number of more controversial examples, and the following principles (which are incompatible with each other) have been proposed:

- Intervocalic consonants go with the following vowel, wherever possible. This is known as the Maximum Onset Principle.
- Consonants go with whichever of the two vowels is more strongly stressed (or, if they are equally stressed, with the preceding vowel).
- Stressed syllables cannot end with a short vowel (i.e., they must be closed with a final consonant; see below) and two consonants are split between the two syllables.
- Allophonic detail may be a strong clue. A clear /l/ between two vowels is perceived as initial in the second syllable, because clear /l/s appear in the onset position. An aspirated plosive is perceived as initial in the second syllable because aspirated plosives appear in the onset position.
- Spelling has been claimed to have some effect on syllabification judgments. In a word like *yellow*, the /l/ is taken to belong to the first syllable because the spelling sequence *ll* cannot start words in English (with the possible exception of foreign words such as *llama*), but can end words such as *yell*.

Eddington, Treiman, and Elzinga (2013a) report that "80% or more of the subjects agreed on the syllabification of 45% of the items with four medial consonants, 69% of the items with three consonants, and 80% of the words with two consonants. What is surprising is that this number drops to 50% for words with a single

medial consonant in spite of the fact that only two syllabification responses are possible." This leaves a fair amount of listener variability, even for examples with only two possible responses, which are split perfectly. Perhaps for this reason, an ambisyllabicity principle has long been proposed whereby an intervocalic consonant can be analyzed as belonging to both the preceding and the following syllable (Anderson and Jones 1974; Lass 1984: 266). By this analysis, the /b/ of a word like *habit* is shared between the two syllables: /[hæ [b]ɪt]/, where the square brackets show syllable boundaries.

For many purposes, these are problems that do not need to be solved. As Wells (2000: xxi), who uses spaces in the *Longman Pronunciation Dictionary* to represent syllable boundaries, says, "any user of the dictionary who finds it difficult to accept the LPD approach to syllabification can simply ignore the syllable spaces."

Semi-vowels, syllabic consonants

In the above explanation, we stated that the onset and coda positions are occupied by consonant sounds and the peak position by vowel sounds. That is not the whole truth and counterexamples now need to be discussed.

Semi-vowels

There are many one-syllable words that have the structure /*ɛt/, that is, their peak and coda (rhyme) is /ɛt/. They include *pet*, *bet*, *debt*, *get*, *jet*, *vet*, *set*, *met*, *net*, *het (up)*. These all clearly have /ɛt/ preceded by a consonant sound (/pɛt, bɛt, dɛt, gɛt, dʒɛt, vɛt, sɛt, mɛt, nɛt, hɛt/).

There are also the words *yet* and *wet*, although it may be unclear whether the initial sounds are consonants or not. In answering this, we need to distinguish between phonetic form (the way these sounds are articulated) and phonological function (the way they function in syllables). In terms of function, these sounds seem to occur in the onset position and the words have the same structure /*ɛt/. However, in terms of form, they are unlike the other consonants. If you slow down the initial sounds of *yet* and *wet*, you will appreciate that they are articulated like, and sound like, the vowels /i:/ and /u:/, as in *tea* and *two*. That is, the tongue and lips do not form any substantial obstruction to the airstream, which escapes relatively freely. Therefore, in terms of function, they are nonsyllabic, in that they do not occur in the peak position, but in terms of form they are vowel-like (vocoid). As a result, /j, w/ are often referred to as semi-vowels.

Syllabic consonants

A further complication relates to the pronunciation of words such as *sudden* and *middle*. Both words are clearly two syllables, and in their fullest form would be pronounced /sʌdən, mɪdəl/, that is, with a schwa vowel after the /d/. However,

Table 5.3 Sounds analysed in terms of their phonetic form and phonological function in the syllable.

		Phonological function	
		Syllabic	*Non-syllabic*
Phonetic	Vocoid	Vowels /iː, æ, uː, ɔɪ/ etc	Semi-vowels /j, w/
form	Contoid	Syllabic consonants /n̩, l̩/	Consonants /p, l, n, k/ etc

this pronunciation was described as "strikingly unusual – and even childish" by Gimson (1980: 320), and it is much more usual to run from the /d/ straight into the /n, l/ without any intervening schwa vowel.

Let us analyse this from the articulatory point of view. The /d/ sound is a voiced alveolar oral-stop (plosive). This means that the tongue comes into contact with the alveolar ridge behind the upper teeth, completely stopping the airstream. The /n/ sound is a voiced alveolar nasal-stop. Thus, in going from a /d/ to an /n/, the tongue does not move, as it is already in the required position. Instead, the soft palate (velum) leading to the nose opens, so that air escapes through the nose. This is known as nasal release of the /d/.

A similar transition occurs with /l/, a voiced alveolar lateral-approximant. Therefore, in going from an /d/ to an /l/, the velum does not move, as both sounds are oral. Instead, the tongue sides lose contact, allowing air to escape over the sides; the tongue tip maintains contact. This is known as lateral release of the /d/.

In terms of syllable structure, these pronunciations mean that we have two-syllable words, but with no vowel sound in the second syllable, as the /n, l/ sounds are clearly pronounced like consonants with substantial obstruction to the airstream (contoids). We thus analyse the consonants /n, l/ as occupying the peak position in the second syllable, and label them syllabic consonants. They are shown by a subscript tick: /sʌdn̩, mɪdl̩/. This situation is summarized in Table 5.3.

Some syllable structure rules of English

It is, of course, impossible to discuss all the permutations of phonemes allowed by the syllable structure of English in any depth in this chapter. For a more thorough description, see Cruttenden (2008: sec. 10.10) for BrE and Kreidler (2008: chs 5 and 6) for AmE. Instead, a few selected generalizations about English syllable structure will be examined. The first is designed to elicit various problems with analyzing the syllable structure of English.

/s/ + consonant initial clusters

A common pattern for two-consonant initial clusters is for the first consonant to be /s/. Although there are 24 consonants in English, fewer than half of them can follow /s/ in a CC cluster. Consonants cannot follow themselves in clusters, that is, there is no /ss/ initial cluster. Those that can follow /s/ fall into categories by manner of articulation (the kind of sound they are):

- Oral-stops (plosives): All three voiceless oral-stops of English can follow /s/, as in *span*, *stuck*, *skill* /spæn, stʌk, skɪl/. Perceptive readers may wonder why the clusters /sb, sd, sg/ do not occur or, more profoundly, why the clusters in *span*, *stuck*, *skill* are analyzed as /sp, st, sk/ rather than /sb, sd, sg/, that is, the voiced equivalents. The answer is that they could as easily be analyzed as /sb, sd, sg/ (and identical clusters in Italian are analyzed in this way). The sounds following /s/ are (i) voiceless (as the voiceless /p, t, k/ in *pan*, *tuck*, *kill* are) but (ii) unaspirated, that is, there is no burst of voiceless air when the sound is released (as is true of /b, d, g/ in *ban*, *duck*, *gill*). In other words, the sounds resemble both /p, t, k/ and /b, d, g/ and could be analyzed either way. The fact that they are represented by *p*, *t*, *k* in English spelling may influence the analysis here. (Similarly, for word-medial sequences, see Davidsen-Nielsen (1974), who found that listeners could not distinguish *disperse* from *disburse*.)
- Nasal-stops: /s/ can be followed by /m, n/ as in *small*, *snow*. It cannot be followed by the third nasal-stop of English because /ŋ/ never occurs in the onset position (see above).
- Fricatives: /s/ can be followed by /f/, as in *sphere*, and /v/, as in *svelte*. However, there is clearly something odd about these clusters.
  - The /sf/ cluster only occurs in a handful of words in English: *sphagnum* (moss), *sphincter*, *sphinx*. Furthermore, all these words (and many other technical words with /sf/) are of Greek origin; /sf/ is a regular cluster in Greek. For these and other reasons (see below), we may consider /sf/ irregular in English.
  - Similarly, the /sv/ cluster only occurs in the one word *svelte* (and *svengali*?), and even there it may be pronounced with /sf/. This word is from Latin, via Italian and French. The name *Sven* is Swedish and may be pronounced with /sv/ in English. However, it is often regularized to /sw/; many Swedes regularized their surname from Svensen to Swensen when they migrated to the USA, including Earle Swensen, the founder of the restaurant chain.
- Approximants: There are four approximants in English: /l, r, w, j/. Of these, two are uncontroversial: /sl/ as in *sleep* and /sw/ as in *swim*. However, /r, j/ need further discussion.
  - The cluster /sr/ only occurs in *Sri Lanka* (and other clearly foreign words). It may therefore be considered non-English. Indeed, many speakers pronounce this country with /ʃri:/, that is, making the initial cluster regular, as in *shrimp*, etc.

- The cluster /sj/ may occur for some BrE speakers in words like *suit* and *assume*. However, this cluster seems to be getting rarer, being replaced by one of two things. Firstly, the /j/ may disappear, leaving plain /s/, this pronunciation becoming increasingly common in British accents. Wells (2008: 790) gives a graph showing that the pronunciation without /j/ is almost exclusively preferred by younger speakers, whereas older speakers often favored /sju:t/. The second possibility is for the (underlying) /sj/ to coalesce into /ʃ/ (see Chapter 9 on connected speech processes). This is quite common in some accents with words like *assume* (thus /əʃu:m/). Neither of these possibilities exists in AmE /su:t, əsu:m/.

In summary, only seven consonants follow initial /s/ uncontroversially, while another four are dubious.

Plosive + approximant initial clusters

If all the permutations of the six plosives /p, b, t, d, k, g/ and the four approximants /l, r, w, j/ existed, there would be 24 possible combinations. However, only 18 of the 24 possible combinations occur, e.g., *play*, *bring*, *quick* /pleɪ, brɪŋ, kwɪk/. The following do not occur: /pw, bw, tl, dl, gw, gj/. There are some rare words and foreign loanwords that contain these clusters (e.g., *pueblo*, *bwana*, *Tlingit*, *guava*, *guano*, *gules*), but no common native words.

Three-consonant initial clusters can be considered a combination of the two patterns just described. In such clusters in English, the first consonant can only be /s/, the second must be a voiceless plosive /p, t, k/, and the third an approximant /l, r, w, j/, e.g., *spring* /sprɪŋ/, *split* /splɪt/, *squid* /skwɪd/. However, again, not all 12 possible permutations occur: /spw, stl, stw/ do not exist.

Rhoticity

Some speakers of English pronounce two /r/ sounds in the phrase *car park*, while others pronounce none. That is, speakers either have both or neither of what is represented by the *r* letter in the spelling. In phonological terms, speakers either can or cannot have /r/ in the coda position in the syllable. Accents that have syllable-final /r/ are called *rhotic*, while the others are *nonrhotic*.

This difference is pervasive throughout the phonology of accents of English. In Shakespeare's day, all speakers of English had syllable-final /r/ (were rhotic). However, a change spread from the Southeast of England and /r/ was dropped in the coda position. This nonrhoticity change spread to most areas of England and Wales; however, it did not affect Scotland and Ireland. The status of countries that England colonized or where native speakers migrated depends on the most influential part of Britain that they came from. Nonrhotic accents include Australia, New Zealand, South Africa, Trinidad, certain eastern and southern parts of the United States, and most of England and Wales. Rhotic accents include Scotland, Ireland, Canada, Barbados, certain western parts of England, and most of the

United States. Because most US speakers are rhotic, rhotic speakers are in the majority in global numerical terms.

Learners of English tend to be rhotic if (i) their native language allows syllable-final /r/ and/or (ii) AmE is more influential than BrE.

The /ʒ/ phoneme

The phoneme /ʒ/ is a peculiar one in the phonology of English. It commonly occurs in the middle of words such as *vision*, as the result of a historical process like that described above for *assume*. It is a moot point whether this is final in the first syllable or initial in the second. On occasions like this, analysts often consider whether the sound(s) can occur at the beginnings or ends of words, that is, as the onset of the first syllable or the coda of the final syllable of a multisyllable word. However, in this case this is inconclusive as no native English words begin or end with /ʒ/. There are plenty of words used in English that begin or end with /ʒ/, but they all have clear foreign origins, usually French: *gendarme*, *genre*, *Giselle*, *je ne sais quoi*, *joie de vivre*; *barrage*, *beige*, *blancmange*, *camouflage*, *collage*, *cortege*, *dressage*, *entourage*, *espionage*, *fuselage*, *liege*, *luge*, *massage*, *mirage*, *montage*, *prestige*, *rouge*.

One of three things can happen with loanwords like these:

1. If English speakers know French, and can pronounce initial or final /ʒ/, then they may pronounce it with that sound.
2. If English speakers do not know French, and/or cannot pronounce initial or final /ʒ/, then they may substitute the closest native English sound, which is /dʒ/. In this way, *beige* (a French loanword) and *page* (a native English word) rhyme.
3. Words may be more fully integrated into English phonology. This is the case with *garage*. By the first process described above, this is /gærɑːʒ/. By the second, it is /gærɑːdʒ/. In AmE, French loanwords are typically pronounced with stress on the second syllable, e.g., *ballet*, AmE /bæˈleɪ/, BrE /ˈbæleɪ/. Similarly, *garage* typically has stress on the second syllable in AmE, but on the first in BrE. A possible BrE pronunciation is thus /ˈgærɑːdʒ/. The weakening of unstressed syllables, very common in BrE, then changes the /ɑː/ vowel to /ɪ/, giving /ˈgærɪdʒ/, which rhymes with *marriage*.

Notice, however, that the above processes depend on how recently the word was borrowed into English and whether it has been fully integrated, like *garage*. Other fully integrated French loanwords include *mortgage* and *visage*.

Open syllables

One-syllable words that have no final consonant sound fall into only two categories, in terms of the vowel:

- Long monophthong vowels: /iː/ (*see* /siː/), /ɑː/ (*shah* /ʃɑː/), /ɔː/ (*law* /lɔː/), /uː/ (*shoe* /ʃuː/) (and /ɜː/ (*fur* /fɜː/) in nonrhotic BrE).

- Diphthong vowels (which are also long): /eɪ/ (*way* /weɪ/), /aɪ/ (*fly* /flaɪ/), /ɔɪ/ (*boy* /bɔɪ/), /aʊ/ (*cow* /kaʊ/), /oʊ/ (*go* /goʊ/), (and /ɪə/ (*here* /hɪə/), /eə/ (*there* /ðeə/), /ʊə/ (*pure* /pjʊə/) in nonrhotic BrE).

Such syllables, without final consonants, that is, with empty codas, are termed open.

Final /ŋ/ sound

Words that contain a final /ŋ/ sound fall into only one category, in terms of the vowel preceding the /ŋ/. The vowel preceding the /ŋ/ is a short monophthong vowel: /ɪ/ (*ring* /rɪŋ/), /æ/ (*hang* /hæŋ/), /ʌ/ (*tongue* /tʌŋ/). Examples with the other short vowels are rarer, e.g., *length* /lɛŋθ/, *kung fu* /kʊŋ fuː/. BrE examples with /ɒ/ (*long* /lɒŋ/) continue this pattern, although they are pronounced with long /ɑː/ in AmE.

In summary, only long vowels can occur in open syllables with no final consonant. Secondly, short vowels can only occur in closed syllables with a final consonant. This is true with /ŋ/; the only exceptions are examples where assimilation (see Chapter 9 on connected speech processes) has taken place, e.g., *green card* /griːn kɑː(r)d/ > /griːŋ kɑː(r)d/. It is also true with other consonants, e.g., *tip, bet, sack, bomb, good* /tɪp, bɛt, sæk, bɒm, gʊd/ occur, but not /tɪ, bɛ, sæ, (bɒ,) gʊ/.

Consonant + /j/ initial clusters

There are plenty of examples of consonant + /j/ initial clusters: *puma* /pjuːmə/, *cute* /kjuːt/, *future* /fjuːtʃər/, *music* /mjuːzɪk/. The generalization here is that the vowel that follows a CC initial cluster, where /j/ is the second consonant, must be /uː/ (or /ʊə/, especially before /r/ in BrE, e.g., *puerile* /pjʊəraɪl/, *cure* /kjʊə/, *furious* /fjʊərɪəs/, *mural* /mjʊərəl/). In unstressed syllables, this can weaken to /ʊ/ or /ə/, e.g., *regular* /rɛgjʊlə, rɛgjələ/.

In AmE, /j/ does not occur in clusters after dental and alveolar sounds, e.g., *enthusiasm*, *tune*, *news* AmE /ɪnθuːzɪæzəm, tuːn, nuːz/, BrE /ɪnθjuːzɪæzəm, tjuːn, njuːz/.

Three- and four-consonant syllable-final clusters

Syllable-final clusters can contain up to three or even four consonants. However, the consonant phonemes that can function as the third or fourth consonant of such clusters are very limited. They are /t, d, s, z, θ/. The list is limited because very often these represent suffixes:

- Past tense verbs or participles, e.g., *lapse* > *lapsed*
- Plural nouns, e.g., *lamp* > *lamps*
- Third person singular present tense verbs, e.g., *ask* > *asks*
- Possessive nouns, e.g., *student* > *student's*
- Contractions of *is* or *has*, e.g., *The bank's opening*, *The bank's opened*

- Ordinal numbers, e.g., *twelve* > *twelfth*
- Quality nouns, e.g., *warm* > *warmth*

In short, English is a language that makes extensive use of inflexions, derivations, and contractions, many of which contribute to the size of syllable-final clusters. Syllable structure is thus connected with grammar and morphology here.

It is worthwhile remembering that languages change over time and that the syllable structure rules in this paper are those of modern-day English speakers. For example, the spelling of *gnat* and *knight* contains vestigial *g* and *k* letters, because these words used to be pronounced /gn, kn/, as some people still do with *gnu* (and compare with German *Knecht*).

Potential syllables

The syllable structure "rules" that we have discussed above are only rules in the sense that they are generalizations about what does and does not occur in English. As we have seen, exceptions to the rules may occur, for instance, because of loanwords that have not been fully integrated.

A pertinent question is whether the syllable structure rules are better than a simple list of all the syllables that occur in English words. They are better, because they describe patterns from a phonological viewpoint.

This may be illustrated by considering the syllables /spɪə, slɪə, sθɪə, sfɪə/ (AmE /spɪr, slɪr, sθɪr, sfɪr/). We will analyse them by considering two factors: (i) whether they are regular, that is, they follow the rules and (ii) whether they are occurring, that is, they exist as or in words of English.

- ***Regular and occurring:*** The syllable /spɪə/ is both regular and occurring – it is the word *spear*. Since the syllable structure rules are generalizations based on the vocabulary of English, it is not surprising that (almost all) occurring syllables are also regular.
- ***Regular but not occurring:*** The syllable /slɪə/ is also regular in that it does not break any of the syllable structure rules of English. The cluster /sl/ is a permissible initial cluster, as in *sleep*. The /l/ consonant can be followed by the /ɪə/ vowel, as in *leer*. However, /slɪə/ happens not to occur as a word of English or as a syllable in a multisyllable word. We can call this a *potential* syllable of English.
- ***Irregular and not occurring:*** The syllable /sθɪə/ is not occurring. It is also not regular, because /sθ/ is not a permissible initial cluster in English. There are, for example, no words beginning /sθ/. In short, the initial cluster /sθ/, and therefore the whole syllable /sθɪə/, does not sound English. One could not imagine naming a new commercial product /sθɪə/, while one might name it /slɪə/.
- ***Irregular but occurring:*** This combination may seem paradoxical, given that we have said that the syllable structure rules are based on the vocabulary of

English. However, syllables and words of this type are usually either of foreign origin or onomatopoeic. We have already mentioned the syllable /sfɪə/, which is occurring, because it is the word *sphere*. However, it is irregular for the reasons given above and because:

- There are no other two-consonant initial clusters in English where the two consonants have the same manner of articulation, i.e., two plosives such as /kt/, two nasals such as /mn/, etc., with the possible exception of /lj/ as in *lurid*.
- There are no consistent patterns of two-consonant initial clusters in English where the second consonant is a fricative.

Onomatopoeic examples include *oink* and *boing*, in both of which the sounds represent the noise of the object (a pig and a spring). However, both break the rule examined above that long vowels (including diphthongs) do not occur before final /ŋ/. In short, there are fuzzy edges to many of the rules of English syllable structure.

Integration of loanwords

Ralph Waldo Emerson, the nineteenth century American essayist, described English as "the sea which receives tributaries from every region under heaven". In other words, English has borrowed words from many languages with which it has come into contact, often through colonization.

The question in this section is, "How are loanwords treated when they are borrowed?" That is, are they integrated into the phonology of the borrowing language or are they left in the same form (phonological and/or orthographic) as in the lending language?

Some languages regularly integrate loanwords into their phonological system. This integration may take different forms:

- Where the loanword contains a sound not present in the borrowing language's segmental inventory, the closest sound is usually substituted. For instance, the word *loch*, as in Loch Ness, is from Scottish Gaelic. The final sound is a voiceless velar fricative /x/. Since this is not a native English sound, the voiceless velar plosive /k/ is often substituted (much to the annoyance of the Scots).
- Where the loanword contains clusters that are not permissible in the borrowing language, these clusters can be broken up by the insertion of vowels. For example, Japanese allows no clusters; its syllables are mostly CVO, with /n/ being the only permissible final consonant. When a word like *screwdriver*, with its /skr/ and /dr/ clusters, is borrowed into Japanese, it is pronounced /sikorudoraiba/. Note that, because vowels have been added, there are now six syllables in the Japanese pronunciation, compared with just three in the English. Note also that Japanese has no /v/ sound and /b/ has been substituted.

An alternative method of dealing with clusters is for a language to simplify them by omitting one or more constituent sound. For instance, the word *cent* was borrowed into Malay, which does not allow final clusters, as /sen/.

- Where the loanword contains final consonants that are not permissible in the borrowing language, these may be simply omitted. An alternative is for a vowel to be inserted after the consonant, effectively making a new syllable with the original final consonant as its initial consonant. For instance, Māori is a CVO language and allows no final consonants; its syllable structure is very similar to Japanese in this respect. It also has a small consonant inventory: /p, t, k, f, h, m, n, ŋ, r, w/. When Westerners brought concepts like the *sheep*, *bus*, *snake*, *biscuit*, and *football* to New Zealand, these loanwords were integrated into Māori as /hipi, pahi, neke, pihikete, futuporo/. Note, among other things, that a vowel has been added after the final consonant in the English words.

English has tended not to integrate loanwords. The example of *sphere* was given above; when it was borrowed into English from Greek, the /sf/ initial cluster was not changed (for example, to /sp/) to conform to English syllable structure rules. However, English has integrated some loanwords. For instance, the German word *Schnorchel* was borrowed as the English word *snorkel*. Note that (i) the /ʃn/ cluster is impermissible in English words (apart from other borrowings such as the German *schnapps*) and has been changed to the native /sn/ cluster, (ii) the voiceless velar fricative [x] in the German pronunciation does not occur in English and is substituted by the closest native sound, the voiceless velar plosive /k/, and (iii) the spelling has been changed to reflect these changes in the pronunciation.

Syllables in pronunciation teaching

As was noted above, syllable structure is important not only in pronunciation teaching but also in literacy. Like literacy experts, pronunciation teachers maintain that there are several features of the syllable that need to be mastered.

Number of syllables

Learners need to be able to say how many syllables multisyllable words contain. Several books (e.g., Gilbert 2001: 12–18) contain exercises in stating how many syllables words contain and in tapping out the syllables.

Separating the onset and rhyme

We have seen how the onset position functions somewhat independently of the rhyme (peak and coda positions). The phenomena of alliteration and rhyme relate to these two components respectively. Activities such as those in Vaughan-Rees (2010) can be used to reinforce these features. Questions such as the following can be asked: "Which word does not rhyme: *spoon*, *book*, *tune*?", "Which word has a different first sound: *chair*, *call*, *kick*?"

In terms of literacy, the spelling of the onset is largely independent of the rhyme. Activities can be used that highlight this. For instance, learners can be asked to put the consonants /b, kl, d, g, h, dʒ, m, p, s, st, sl, θ/ before the rhyme /ʌmp/ or, in spelling terms, the letters *b*, *cl*, *d*, *g*, *h*, *j*, *m*, *p*, *s*, *st*, *sl*, *th* before *ump*.

Separating the peak and coda

Similar activities can be used for distinguishing the vowel in the peak from the consonant(s) in the coda. In short, the three positions and their constituent sounds can be worked on independently or in various combinations: "Here is a picture of a bell. Finish the word for me: /bɛ … /", "Say *spoon*. Now say it again, but instead of /u:/ say /ɪ/", "Say *trip*. Now say it without the /r/ sound", "Say *pink*. Now say it again but do not say /p/."

Dealing with clusters

Consonant clusters are a major problem, especially for learners from languages that do not permit clusters. Various (combinations of) strategies, similar to those described above for the integration of loanwords, may be resorted to by the learner. The following examples relate to the English word *street* /stri:t/.

- Final consonants may simply be omitted: /stri:/.
- Extra vowels may be added to final consonants: /stri:ti:/.
- Extra vowels may be inserted to break up initial clusters: /si:tri:t/.
- Extra vowels may be inserted before initial clusters: /i:stri:t/.

The fact should not be forgotten that words are not said in isolation, but in stretches that are linked together (see Chapter 9 on connected speech processes). Thus, if learners omit the final /k/ of *link*, they should be given practice in pronouncing the word with something following that begins with a vowel sound, e.g., *link it*, *linking*. The word division is largely irrelevant here; *link it* ends like (rhymes with) *trinket*.

Likewise, linked phrases can be used to combat the above learners' strategies. In *street address*, the final /t/ is linked to the following vowel, avoiding deletion of the /t/ or insertion of a vowel. In *this street*, the fact that *this* ends in the same /s/ consonant as at the beginning of *street* allows the two /s/s to be joined, avoiding intrusive vowels.

Conclusion

Syllables have been shown to be an important, but often overlooked, aspect of the phonology of languages. Many of the features encountered when speakers of other languages learn English, or indeed when speakers of English learn other languages, can be simply explained in terms of the syllable structure possibilities in

the two languages. These problems are especially acute for learners of English, as English has a more complex syllable structure than that of most other languages.

REFERENCES

Abercrombie, D. 1967. *Elements of General Phonetics*, Edinburgh: Edinburgh University Press.

Anderson, G.D.S. n.d. a. The velar nasal. In: *World Atlas of Language Structures Online*, Chapter 9. wals.info/chapter/9.

Anderson, G.D.S. n.d. b. Areal and phonotactic distribution of /ŋ/. www.let.leidenuniv.nl/ulcl/events/ocp1/abstracts/Anderson.pdf.

Anderson, J. and Jones, C. 1974. Three theses concerning phonological representations. *Journal of Linguistics* 10: 1–26.

Bowring, R. and Laurie, H.U. 2004. *An Introduction to Modern Japanese: Book 1*. Cambridge: Cambridge University Press.

Carson, K.L., Gillon, G.T., and Boustead, T.M. 2013. Classroom phonological awareness instruction and literacy outcomes in the first year of school. *Language Speech and Hearing Services in Schools* 44(2): 147–160.

Cruttenden, A. 2008. *Gimson's Pronunciation of English*, 7th edition, London: Edward Arnold.

Cutler, A. (ed.) 1982. *Slips of the Tongue and Language Production*, Berlin: de Gruyter.

Davidsen-Nielsen, N. 1974. Syllabification in English words with medial sp, st, sk clusters. *Journal of Phonetics* 2: 15–45.

Eddington, D., Treiman, R., and Elzinga, D. 2013a. Syllabification of American English: evidence from a large-scale experiment. *Part I. Journal of Quantitative Linguistics* 20(1): 45–67.

Eddington, D., Treiman, R., and Elzinga, D. 2013b. Syllabification of American English: evidence from a large-scale experiment. *Part II. Journal of Quantitative Linguistics* 20(2): 75–93.

Fromkin, V.A. (ed.) 1980. *Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen and Hand*, New York: Academic Press.

Gilbert, J.B. 2001. *Clear Speech from the Start*, Cambridge: Cambridge University Press.

Gimson, A.C. 1980. *An Introduction to the Pronunciation of English*, 3rd edition, London: Edward Arnold.

Gnanadesikan, A.E. 2008. Syllables and syllabaries: what writing systems tell us about syllable structure. In: *CUNY Conference on the Syllable*, City University of New York, 17–19 January. www.cunyphonologyforum.net/SYLLABSTRACTS/Gnanadesikan.pdf.

HearBuilder. n.d. *Phonological Awareness* (software). www.hearbuilder.com/phonologicalawareness.

Justice, L.M., Gillon, G.T., McNeill, B.C., and Scheule, C.M. 2013. Phonological awareness: description, assessment, and intervention. In: *Articulation and Phonological Disorders: Speech Sound Disorders in Children*, J.E. Bernthal, N.W. Bankson, and P. Flipsen (eds.), 7th edition, 355–382, Boston: Pearson.

Kreidler, C.W. 2008. *The Pronunciation of English: A Course Book*, 2nd edition, Malden, MA: Wiley-Blackwell.

Ladefoged, P. and Johnson, K. 2010. *A Course in Phonetics*, 6th edition, Boston: Wadsworth/Cengage Learning.

Lass, R. 1984. *Phonology: An Introduction to Basic Concepts*, Cambridge: Cambridge University Press.

Maddieson, I. n.d. Syllable structure. In: *World Atlas of Language Structures Online*, Chapter 12. wals.info/chapter/12.

Moats, L. 2010. *Speech to Print: Language Essentials for Teachers*, 2nd edition, Baltimore, MA: Brookes Publishing.

Moats, L. and Tolman, C. 2009. *Language Essentials for Teachers of Reading and Spelling (LETRS)*, Boston: Sopris West.

Redford, M.A. and Randall, P. 2005. The role of juncture cues and phonological knowledge in English syllabification judgments. *Journal of Phonetics* 33: 27–46.

Vaughan-Rees, M. 2010. *Rhymes and Rhythm: A Poem-based Course for English Pronunciation*, Reading: Garnet.

Wells, J.C. 1990. Syllabification and allophony. In: *Studies in the Pronunciation of English*, S. Ramsaran (ed.), 76–86, London: Edward Arnold.

Wells, J.C. 2000. *Longman Pronunciation Dictionary*, 2nd edition, London: Longman.

Wells, J.C. 2008. *Longman Pronunciation Dictionary*, 3rd edition, London: Longman.

Williamson, G. 2010. *Phonological Processes: Natural Ways of Simplifying Speech Production*, Speech Therapy Information and Resources. www.kalgoo.com/15-phonological-processes.html.

Wilson, J. 2013. *Phonological Awareness Training (PAT)*. www.ucl.ac.uk/educational-psychology/cpd/pat.htm.

6 Lexical Stress in English Pronunciation

ANNE CUTLER

English lexical stress and its pronunciation implications

Not all languages have stress and not all languages that do have stress are alike. English is a lexical stress language, which means that in any English word with more than one syllable, the syllables will differ in their relative salience. Some syllables may serve as the locus for prominence-lending accents. Others can never be accented.

In the word *language*, for example, the first syllable is stressed: *LANGuage* (henceforth, upper case will denote a stressed syllable). If the word *language* receives a principal accent in a sentence, either by default (*She studies languages*) or to express contrast (*Did you say language games or anguish games*?), the expression of this accent will be on *language*'s first syllable. The second syllable of *language* is not a permissible location for such accentuation. Even if we contrive a case in which the second syllable by itself is involved in a contrast (*What was the new password again: "language" or "languish"*?), it is more natural to express this contrast by lengthening the final affricate/fricative rather than by making each second syllable stronger than the first. The stress pattern of an English polysyllabic word is as intrinsic to its phonological identity as the string of segments that make it up.[1]

This type of asymmetry across syllables distinguishes stress languages from languages that have no stress in their word phonology (such as, for instance, many Asian languages). Within stress languages, being a lexical stress language means that stress can vary across syllable positions within words, and in principle can vary contrastively; this distinguishes lexical stress languages from fixed-stress languages (such as Polish or Finnish), where stress is assigned to the same syllable position in any word (the penultimate syllable in Polish; the initial syllable in Finnish).

The Handbook of English Pronunciation, First Edition. Edited by Marnie Reed and John M. Levis.

The "in principle" qualification on contrastivity holds not only for English; in all lexical stress languages, minimal pairs of words varying only in stress are rare. English has only a few (*INsight* versus *inCITE* and *FOREbear* versus *forBEAR*, for example); they require two successive syllables with full vowels, and this is in any case rare among English words. Stress alone is not a major source of inter-word contrast in English.

One way in which English does vary stress across words, however, is by the role stress plays in derivational morphology. Adding a derivational affix to an English word, and thus creating a morphologically related word of a different grammatical class, very often moves the location of the primary stress to a different syllable; we can *adMIRE* a *BAron* as a *PERson* who is *aristoCRATic* or express our *admiRAtion* for his *baRONial* ability to *perSONify* the *arisTOCracy*.

Rhythmically, English prefers to avoid successive stressed syllables, and alternation of stressed and unstressed syllables characterizes English speech. There is an obvious implication of this preference for stress alternation, together with the fact that English words may have only one primary stressed syllable but may have three, four, or more syllables in all: there are different levels of stress. Thus in *admiration* and *aristocracy*, with primary stress in each case on the third syllable, the first syllable bears a lesser level of stress (often referred to as "secondary stress"; see the metrical phonology literature, from Liberman and Prince 1977 on, for detailed analyses of relative prominence in English utterances).

Finally, English differs from some other lexical stress languages in how stress is realized. The salience difference between stressed and unstressed syllables is realized in several dimensions; stressed syllables are longer, can be louder and higher in pitch or containing more pitch movement than unstressed syllables, and the distribution of energy across the frequency spectrum may also differ, with more energy in higher-frequency regions in stressed syllables (for the classic references reporting these analyses, see Cutler 2005a). The difference between a stressed and an unstressed version of the same syllable can be clearly seen in Figure 6.1.

All these dimensions are suprasegmental, in that a given sequence of segments retains its segmental identity though it can be uttered in a shorter or longer realization, with higher or lower pitch, and so on (see Lehiste 1970 for a still unsurpassed account of suprasegmental dimensions). All lexical stress languages use such suprasegmental distinctions, but English also distinguishes stressed and unstressed syllables segmentally, in the patterning of vowels. In English, vowels may be full or reduced. Full vowels may be monophthongs (e.g., the vowels in *Al*, *ill*, *eel*) or diphthongs (as in *aisle*, *oil*, *owl*), but they all have full vowel quality. Reduced vowels are centralized, with schwa the most common such vowel (the second vowel in *Alan* or the first in *alone*). Any stressed syllable in English must contain a full vowel (e.g., the first vowel in *language*). Any syllable with a reduced vowel (e.g., *language*'s second syllable) may not bear stress.

In this last feature, English obviously differs from lexical stress languages without reduced vowels in their phonology (e.g., Spanish); in such languages, suprasegmental distinctions are the only means available for marking stress. In English, the segmental reflection of stress is so important that linguists have

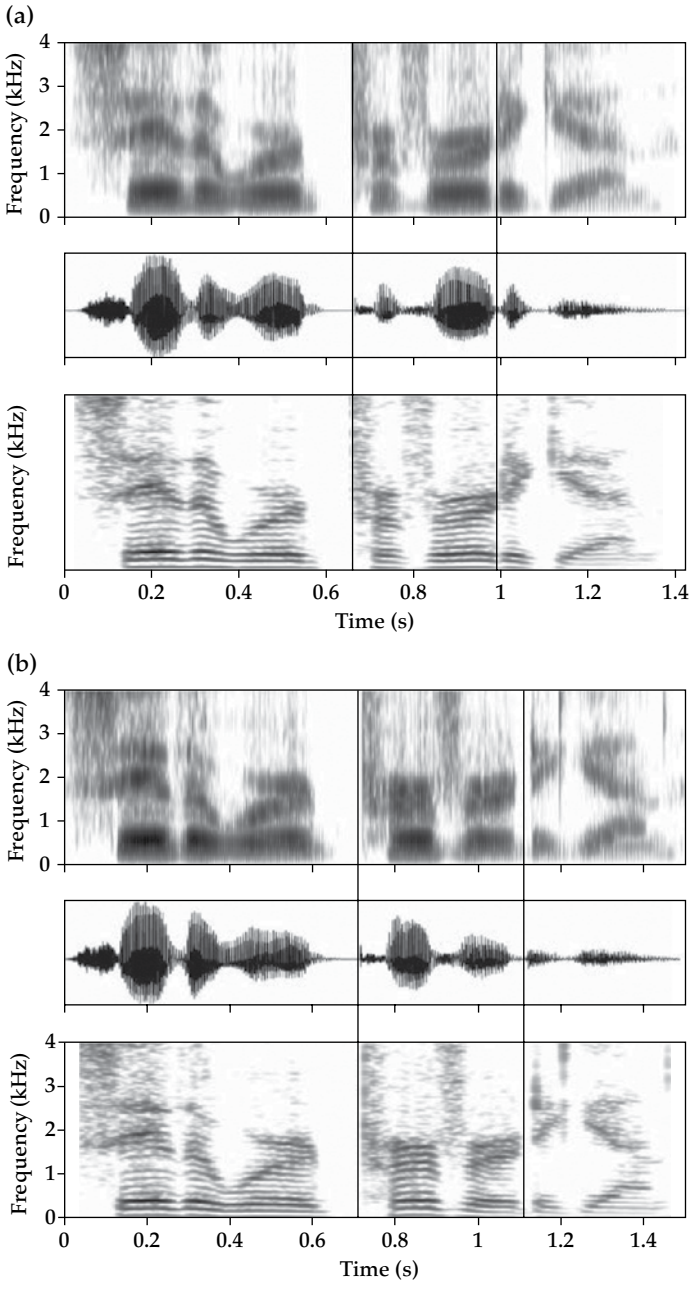


Figure 6.1 The verb *perVERT* (upper three panels) and the noun *PERvert* (lower three panels), which differ in stress, spoken by a male speaker of American English in the carrier sentence *Say the word .... again*. The three display panels of each figure are: top, a broad-band spectrogram; middle, a waveform display; below, a narrow-band spectrogram. Vertical lines in each panel indicate the onset and offset of the example word *pervert*. The figure is modelled on a figure created by Lehiste and Peterson (1959: 434). The stressed syllables (the second syllable of the verb, in the upper panels, and the first syllable of the noun, in the lower panels) are longer, louder, and higher in pitch than the unstressed versions of the same syllables (the first syllable of the verb, the second syllable of the noun). The length difference can be particularly well seen in the broad-band spectrogram, the loudness difference in the waveform, and the pitch difference in the narrow-band spectrogram, where the higher the fundamental frequency (pitch), the wider the spacing of its resonants (the formants, forming stripes in the figure).

observed that it is possible to regard English as a two-level prominence system: full vowels on one level, reduced vowels on the other (Bolinger 1981; Ladefoged 2006). This segmental feature is crucial to the functioning of stress, not only in the phonology but also in language users' production and perception of words and sentences. As we shall see, its role in speech perception in particular entails that when a slip of the tongue or a non-native mispronunciation causes alteration of the patterning of full and reduced vowels, then recognition of the intended word is seriously hindered.

The perception of English lexical stress by native listeners

If lexical stress by itself rarely makes a crucial distinction between words, how important is it for recognizing words? The segmental building blocks of speech – vowels and consonants – certainly do distinguish minimal pairs of words. We need to identify all the sounds of *creek* to be sure that it is not *freak*, *Greek*, *clique*, *croak*, *crack*, *creep*, or *crease*. However, minimal pairs such as *incite/insight* occur so rarely in our listening experience that there would be little cost to the listener in ignoring the stress pattern and treating such pairs as accidental homophones, like *sole/soul*, *rain/rein/reign*, or *medal/meddle*. Languages do not avoid homophony – quite the reverse – in that new meanings tend not to be expressed with totally new phonological forms, but are by preference assigned to existing forms (*web*, *tweet*, *cookies*). This preference occurs across languages and putatively serves the interest of language users by reducing processing effort (Piantadosi, Tily, and Gibson 2012). Indeed, there is evidence from psycholinguistic laboratories that English words with a minimal stress pair do momentarily make the meanings of each member of the pair available in the listener's mind (Cutler 1986; Small, Simon, and Goldberg 1988), just as happens with accidental homophones such as *sole/soul* (Grainger, Van Kang, and Segui 2001).

This by no means implies that stress is ignored by English listeners. The role of any phonological feature in speech perception is determined by its utility; listeners will make use of any speech information if it helps in speech recognition, and they will use it in the way it best helps. Vocabulary analyses show that there is indeed little advantage for English listeners in attending to the suprasegmental reflections of stress pattern over and above the segmental structure of speech, as this achieves only a relatively small reduction in the number of possible words to be considered (Cutler, Norris, and Sebastián-Gallés 2004; Cutler and Pasveer 2006; this English result contrasts significantly with the large reductions achieved when the same analyses are carried out for Spanish, Dutch, and German, all of which are lexical stress languages, but none of which have the strong segmental reflection of stress found in English).

Vocabulary analyses reveal, however, that there is a highly significant tendency for stress in English words to fall on the initial syllable, and this tendency is even greater in real speech samples (Cutler and Carter, 1987).[2] There is an obvious

reason for this: about a quarter of the vocabulary consists of words with unstressed initial syllables, but most of the words in this set have a relatively low frequency of occurrence (*pollution*, *acquire*, *arithmetic*). The higher-frequency words, i.e., the ones most often heard in real speech, are shorter and more likely to have just a single stressed syllable that is either the word-initial syllable (*garbage*, *borrow*, *numbers*) or the only syllable (*trash*, *take*, *math*). This pattern has a very important implication for listeners to English: it means that in any English utterance, a stressed syllable is highly likely to be the beginning of a new word. Since most unstressed syllables are reduced, it is furthermore even a reasonable bet that any syllable containing a full vowel is likely to be the beginning of a new word.

English listeners grasp this probability and act on it. Segmentation of speech signals into their component words is a nontrivial task for listeners, since speech signals are truly continuous – speakers run the words of their utterances together, they do not pause between them. Listeners, however, can only understand utterances by identifying the words that make them up, since many utterances are quite novel. Any highly predictive pattern, such as the English distribution of stress, is therefore going to prove quite useful.

Psycholinguistic experiments with a task called word-spotting, in which listeners detect any real word in a spoken nonsense sequence, provided the first demonstration of English listeners' use of the pattern of full and reduced vowels in segmentation. The input in word-spotting consists of sequences such as *obzel crinthish bookving* and the like (in this case, only the third item contains a real word, namely *book*). A word spread over two syllables with a full vowel in each (e.g., *send* in *sendibe* [sɛndɑɪb]) proved very difficult to detect, but if the same word was spread over a syllable with a full vowel followed by a syllable with a reduced vowel (e.g., *send* in *sendeb* [sɛndəb]), it was much easier to spot (Cutler and Norris 1988). Response times were faster in the latter case and miss rates were lower. In the former case, detection of the embedded word is hindered by the following full vowel because it has induced listeners to segment the sequence at the onset of the second syllable (*sen - dibe*). They act on the strategy described above: any syllable with a full vowel is likely to be a new word. Consequently, detection of *send* requires that its components (*sen*, *d*) be reassembled across this segmentation point. No such delay affects detection of *send* in *sendeb* because no segmentation occurs before a syllable that has a reduced vowel.

Missegmentations of speech show exactly the same pattern. Listeners are far more likely to erroneously assume a stressed syllable to be word-initial and unstressed syllables to be word-internal than vice versa (Cutler and Butterfield 1992). In an experiment with very faint speech, unpredictable sequences such as *conduct ascents uphill* were reported, for instance, as *the doctor sends her bill* – every stressed syllable becoming word-initial here. In collections of natural slips of the ear the same pattern can be observed; thus the song line *she's a must to avoid* was widely reported in the 1980s to have been heard as *she's a muscular boy*, with the stressed last syllable taken as a new word, while the unstressed two syllables preceding it are taken as internal to another word. Jokes about misperception also rely on this natural pattern – an old joke, for instance, had a British Army field

telephone communication *Send reinforcements, we're going to advance* perceived as *Send three-and-fourpence, we're going to a dance*. Once again, stressed syllables have been erroneously assumed to be the beginnings of new words.

This segmentation strategy works well for English and more than compensates for the fact that stress distinctions by themselves do not often distinguish between words. In fact the stress-based segmentation used by English listeners falls in line with strategies used for speech segmentation in other languages, which tend to exploit language-particular rhythmic characteristics. In French and Korean, rhythmic patterns (including poetic patterns) are syllable-based and so is listeners' speech segmentation (Mehler et al. 1981; Kim, Davis, and Cutler 2008). In Japanese and Telugu, rhythm (again, poetic rhythm too) is based on the mora, a subsyllabic unit, and speech segmentation is mora-based too (Otake et al. 1993; Murty, Otake, and Cutler 2007). English, with its stress-based poetic forms and stress-based speech segmentation, further confirms the cross-language utility of speech rhythm for segmentation (see also Chapter 7 on rhythmic structure in this volume).

Given the acoustic reflections of stress described above, visible in Figure 6.1, English stressed syllables are, of course, more easily perceptible than unstressed syllables. They are easier to identify out of context than are unstressed syllables (Lieberman 1963) and speech distortions are more likely to be detected in stressed than in unstressed syllables (Cole, Jakimik, and Cooper 1978; Cole and Jakimik 1980; Browman 1978; Bond and Garnes 1980). Nonwords with initial stress can be repeated more rapidly than nonwords with final stress (Vitevitch et al. 1997; note that such nonwords are also rated to be more word-like, again indicating listeners' sensitivity to the vocabulary probabilities).

However, there is a clear bias in how English listeners decide that a syllable is stress-bearing and hence likely to be word-initial; the primary cue is that the syllable contains a full vowel. Fear, Cutler, and Butterfield (1995) presented listeners with tokens of words such as *audience*, *auditorium*, *audition*, *addition*, in which the initial vowels had been exchanged between words. The participants rated cross-splicings among any of the first three of these as insignificantly different from the original, unspliced tokens. Lower ratings were received only by cross-splicings involving an exchange between, for example, the initial vowel of *addition* (which is reduced) and the initial vowel of any of the other three words. This suggests that preserving the degree of stress (primary stress on the first syllable for *audience* and secondary stress for *auditorium*, an unstressed but full vowel for *audition*) is of relatively little importance compared to preserving the vowel quality (full versus reduced).

In other stress languages, suprasegmental cues to stress can be effectively used to distinguish between words. In Dutch, the first two syllables of *OCtopus* "octopus" and *okTOber* "october" differ only suprasegmentally (not in the vowels), and in Spanish, the first two syllables of *PRINcipe* "prince" and *prinCIpio* "beginning" likewise differ only suprasegmentally. In both these languages, auditory presentation of a two-syllable fragment (*princi-*, *octo-*) significantly assisted subsequent recognition of the matching complete word and significantly delayed subsequent recognition of the mismatching complete word – for example, recognition of *principe* was slower after hearing *prinCI-* than after a neutral control

stimulus (Soto-Faraco, Sebastián-Gallés, and Cutler 2001; Donselaar, Koster, and Cutler 2005). This delay is important: it shows that the word mismatching the spoken input had been ruled out on the basis of the suprasegmental stress cues. The delay is not found in English. Actually, directly analogous experiments are impossible in English since the segmental reflections effectively mean that there are no pairs of the right kind in the vocabulary! In English, the second syllables of *octopus* and *october* are different, because the unstressed one – in *octopus* – has a reduced vowel, which thus is quite different from the stressed vowel in *october*'s second syllable. However, in some word pairs, the first two syllables differ not in where the stressed syllable is but in what degree of stress it carries; for instance, *admi-* from *ADmiral* has primary stress on the first syllable, while *admi-* from *admiRAtion* has secondary stress on the first syllable. In the Dutch and Spanish experiments, such fragments had also been used and had duly led to facilitation for match and delay for mismatch. Cooper, Cutler, and Wales (2002) found a different pattern, however, for such pairs of English words; match facilitated recognition but, crucially, mismatch did not inhibit it, showing that here suprasegmental information for stress had not been used to rule out the item it mismatched.

We conclude, then, that for English listeners the most important reflections of their language's stress patterning are the segmental ones. These are drawn on with great efficiency in parsing utterances and recognizing words. The suprasegmental concomitants of the stress variation, in contrast, are to a large degree actually ignored. Direct evidence for this comes from an experiment by Slowiaczek (1991) in which English listeners heard a sentence context (e.g., *The friendly zookeeper fed the old*) followed by a noise representing a stress pattern (cf. *DAdada* or *daDAda*). The listeners then judged whether a spoken word was the correct continuation of the sentence as signaled by the stress pattern. Slowiaczek found that listeners frequently ignored the stress pattern, for instance accepting *gorilla* as the continuation of this sentence, even when the stress pattern had been *DAdada*, or accepting *elephant* when the stress pattern had been *daDAda*. They apparently attended to the meaning only (a contextually unlikely word, such as *analyst*, thus was rejected whether the stress pattern matched it or not).

Slowiaczek (1990) also found that purely suprasegmental mis-stressing of English words (e.g., switching secondary and primary stress, as in *STAMpede* for *stamPEDE*) did not affect how well noise-masked words were recognized. This was fully in line with the earlier studies, which had shown that the stress pattern did not help to discriminate minimal stress pairs (Cutler 1986; Small, Simon, and Goldberg 1988) and that mis-stressing English words did not inhibit recognition if no segmental change but only suprasegmental changes were made (Bond 1981; Bond and Small 1983; Cutler and Clifton 1984; see also the section below on mispronunciation of stress).

The English vocabulary does not offer much processing advantage for attention to suprasegmental information; English listeners, therefore, largely concentrate on the cues that do provide rapid recognition results, i.e., the segmental cues. Because English stress has segmental as well as suprasegmental realizations, and the segmental patterns are systematically related to the location of word

boundaries, attending chiefly to segmental patterns still allows English listeners to use stress information in segmenting utterances into their component words.

The production of English lexical stress by native speakers

The perceptual evidence does not suggest that speakers adjust suprasegmental parameters separately while articulating English, nor that stress is computed on a word-by-word basis during speech production. Rather, the evidence from perception would be compatible with a view of speech production in which the segmental structure of a to-be-articulated word is retrieved from its stored representation in the mental lexicon, and the metrical pattern of the utterance as a whole is mapped as a consequence of the string of selected words. Exactly such a view is proposed by the leading psycholinguistic modelers of speech production (Levelt 1989; Levelt 1992; Shattuck-Hufnagel 1992; Levelt, Roelofs, and Meyer 1999).

Some relevant evidence comes from slips of the tongue: English native speakers do occasionally make slips in which stress is misplaced (Fromkin 1976; Cutler 1980a). However, it seems that such errors may be an unwanted side-effect of the derivational morphology of English! That is, the errors exhibit a very high likelihood of stress being assigned to a syllable that is appropriately stressed in a morphological relative of the target word. Some examples from published collections of stress errors are: *hierARCHy*, *ecoNOmist*, *homogeNEous*, *cerTIFication*. These four words should have received primary stress respectively on their first, second, third, and fourth syllables, but the stress has been misplaced. It has not been randomly misplaced, however; it has landed precisely on the syllable that bears it in the intended words' relatives *hierarchical*, *economics*, *homogeneity*, and *certificate* respectively.

This pattern suggests, firstly, that words with a derivational morphological relationship are stored in proximity to one another in the speakers' mental lexicon. This is certainly as would be expected given that the organization of a production lexicon serves a system in which meaning is activated first, to be encoded via word forms located in the lexical store. Secondly, the stress error facts suggest that the location of primary stress is represented in these stored forms in an abstract way: given the typical patterning of such derivationally related sets of English words, in many cases the mis-stressing led to a vowel change. Again, this makes sense: each word has its canonical segmental structure (sequence of vowels and consonants) represented in the lexicon, and since words may have more than one syllable with a full vowel, an abstract code is needed to indicate which syllable should receive primary stress. In a stress error, the marking assigned to a particular syllable in one word among a group of related entries has accidentally been applied to the same syllable in another word.

In producing an utterance, then, speakers have to construct an overall smooth contour in which each of the selected words is appropriately uttered and, most importantly, in which the meaning of the utterance as a whole (for instance, the

focal emphasis, the expression of a statement or of a question, and the relation of the words in the utterance to the ongoing discourse) is correctly captured. Pitch accents will be applied in accord with the choices driven by such discourse constraints (see Shattuck-Hufnagel and Turk 1996 for much relevant evidence). Remaining in the domain of lexical stress, where the pitch accents fall will be determined by the markings that, within any polysyllabic word, denote the location of primary stress. As already described, only a stressed syllable can be accented in a sentence.

There is considerable evidence that speakers plan a metrical structure for their utterance and that it is based on the alternating rhythm described in the first section above (see, for example, Cummins and Port 1998). English slips of the tongue in which a syllable is accidentally omitted or added tend to lead to a more regular rhythm than the correct utterance would have had (Cutler 1980b), a pattern that is also found in the way syllables are added by optional epenthesis in the rhythmically similar language Dutch (Kuijpers and Donselaar 1998). Experiments in which speakers are asked to read words from a screen or recall arbitrary word pairs have been shown to elicit faster responses when successive words have the same stress pattern (e.g., Roelofs and Meyer 1998 for Dutch and Colombo and Zevin 2009 for Italian); however, careful explorations with such tasks in English by Shaw (2012) have shown that the facilitation – in this language at least – is not due to activation of a stored template of the metrical pattern. Iambic words (*detach*, *lapel*, etc.) were read out more rapidly after any repeating stress sequence (iambic: *belong*, *canal*, *forgive* or trochaic: *reckon*, *salad*, *fidget*) than after any varying sequence (*salad*, *belong*, *reckon* or *salad*, *reckon*, *belong*). Instead, the facilitated production seems to arise here from predictability of a repeating pattern for articulation. This argues against the metrical pattern of a word in an utterance being a template that is stored as a whole in the lexicon; instead, what is stored is, as suggested above, the segmental structure of the word, along with a code marking the position on which primary stress may fall. All other aspects of a word's metrical realization in an utterance fall out of the word's sequence of syllables containing full versus reduced vowels.

Mispronunciation of stress

Although the evidence from slips of the tongue suggests that stress errors will not occur very often (because they tend to involve multisyllabic derivationally complex words with derivationally complex relatives, and such words have a fairly low frequency of occurrence anyway), it is nevertheless interesting to consider what effects mis-stressing would have on the acoustic realization of a word and on how the word is perceived.

The first syllable of any polysyllabic word may be either stressed (with a full vowel) or unstressed (with a reduced vowel). If the correct pronunciation of the initial syllable has a reduced vowel, then a speaker who is mispronouncing has little option but to alter the vowel quality. Mispronouncing any stressed syllable can also involve changing the vowel (either to a reduced vowel or to any other and

hence incorrect vowel). We saw that English listeners do not attend much to suprasegmental cues in recognizing words, but they do pay great attention to the pattern of strong and weak vowel realizations (especially in their lexical segmentation). Thus the kind of mispronunciation that alters vowel quality should be one that is highly likely to impede successful recognition of the word by native listeners, and repeated experimental demonstrations in the 1980s confirmed that this is indeed so. The results include:

- Different kinds of phonetic distortion impact upon word recognition in differing ways, but the most disruptive type of distortion is changing a vowel, and particularly changing a vowel in a stressed syllable (Bond 1981).
- Shadowing (repeating back) incoming speech is only disrupted by mis-stressing if the mis-stressing involves a change in vowel quality (Bond and Small 1983).
- Semantic judgments on spoken words are also relatively unaffected by mis-stressing except when the misplacement leads to a vowel quality change (Cutler and Clifton 1984).
- Any vowel quality change is equally disruptive; the number of distinctive features involved is irrelevant (Small and Squibb 1989).

The reason for this pattern is to be found in how spoken-word recognition works. When a speech signal reaches a listener's ear, the words that are potentially contained in the incoming utterance automatically become available for consideration by the listener's mind – a process known as lexical *activation*. The word "potentially" is important here; frequently it is the case that many more candidate words are fleetingly activated than the utterance actually contains. Consider the utterance: *Many vacant shops were demolished*. These five words present the listeners with a range of such fleeting possibilities: (a) the first word that is fully compatible with the incoming signal is actually *men*; (b) by the second syllable, *many* is also activated, but that second syllable could also combine with the third to make a word beginning *eva-*, i.e., the utterance might be *men evade* …; (c) the sequence of the reduced syllable *-cant* and the syllable *shop* could be *can chop*; (d) assuming that *were* is unstressed, then *were* plus the unstressed initial *de-* of *demolished* is a possible utterance of *would a*; (e) the stressed syllable of *demolished* could briefly activate words beginning with that syllable, such as *molecule*, *mollify*.

We are usually quite unaware of all such potentially present words in the speech we hear, and of their brief activation, as we rapidly and certainly settle on the correct interpretation of an utterance; but decades of research on spoken-word recognition have shown that this is indeed how this efficient process works (for more detail, see the review by McQueen 2007 or the relevant chapters in Cutler 2012). It is a process in which alternative interpretations of the signal compete with one another, in that the more support any one word receives from the signal, the less likely the other interpretations become. If a candidate word is mismatched by the input, the mismatch has immediate effect and the word is no longer a viable choice (in the above example, *men evade* becomes an impossible interpretation once

the /k/ of *many vac-* arrives. Relevant spoken-word recognition evidence may be found in Vitevitch and Luce 1998 and Soto-Faraco, Sebastián-Gallés, and Cutler 2001). Interestingly, the effects of mismatch can be automatically modulated by the listener if background noise suggests that the signal might be unreliable (McQueen and Huettig 2012; Brouwer, Mitterer, and Huettig 2012), but the standard setting is that mismatch instantly counts against mismatched candidates.

Consider therefore what will happen when a word is mispronounced in any way: the input will activate a population of candidate words that may deviate from the set of candidates a correctly pronounced version would have activated. In the worst case, the intended word will not even be included in the activated set.

Obviously the effects of mismatch mean that to keep the intended word in the set as much as possible it must be correct from the beginning, so that the "safest" mispronunciation, so to speak, is one right at the end of a word. This will lead to misrecognition only if the utterance happens to correspond to an existing word – as when speakers of languages with obligatory devoicing mispronounce finally voiced English words that happen to have a finally unvoiced minimal pair (e.g., saying *save* as if it were *safe* or *prize* as if it were *price*). In many or even most cases, however, a final mispronunciation will not lead to misrecognition – the target word will have been recognized before the mispronunciation arrives (*telephome* and *ostridge* and *splendith* are fairly easy to reconstruct despite the final mispronunciations of place of articulation, voicing, and manner respectively). The very same mispronunciations in the word-initial position, in contrast – say, *motable*, *jeeky*, *thrastic* – make the words harder to reconstruct even when we see them in writing with all the word available at once; even then, the wrong beginning throws us off. The spoken form, coming in over time rather than all at once, misleads us even more decisively. In the case of *motable*, the incoming speech signal could initially call up *mow*, *moat*, *motor*; the input *jeeky* may call up *gee*, *jeep*, *jeans*; and *thrastic* may call up *three*, *thread*, *thrash*. That is, the sets of lexical candidates will at first not even include *notable*, *cheeky*, or *drastic*, and the chance of finding them as the intended word depends, firstly, on the eventual realization that none of the activated word candidates actually matches the signal, followed, secondly, by a decision, perhaps by trial and error, that the offending mispronunciation is in the initial phoneme.

Mis-stressing can cause similar difficulty for the listener whenever it affects the segments that make up the word – that is, whenever a vowel is changed. Mis-stressing will NOT cause difficulty if it involves suprasegmentals only, e.g., when secondary and primary stresses are interchanged; as the early research already mentioned has shown, mis-stressed words where vowels are unchanged (e.g., *stampede* pronounced as *STAMpede*) are recognized easily. However, such mis-stressing can only happen in words with two full vowels (like *stampede*), and, though words of this type can be readily found for experimental purposes, there are in fact not so many of them and they do not occur often in real speech. Stress and vowel realization are so tightly interwoven in the English lexicon, and the lexicon is so strongly biased towards short words and towards words with initial stress, that the most common word type in the vocabulary is a bisyllable with a full

vowel in the first syllable and a weak vowel in the second (e.g., *common*, *vowel*, *second*). Real speech actually contains a majority of monosyllables (where the possibility of mis-stressing does not arise), because the shortest words in the vocabulary are the ones that are used most frequently. As described in the section above on the production of English lexical stress by native speakers, the polysyllabic words in real speech conform even more strongly to the preferred patterns than does the vocabulary as a whole. In other words, where there is opportunity for mis-stressing in real speech, it is most likely to involve a word with stress on its initial syllable and a reduced vowel in its unstressed syllable(s). Thus on average any mis-stressing will indeed involve a vowel change and thus be hard for listeners to recognize.

Consider some examples and the consequent activated lexical candidates. Again the rule holds that early effects of mis-stressing are more harmful to recognition than later effects. *Common* with stress shifted to the second syllable and a reduced first vowel could initially activate a large set of words with unstressed initial *com- – commodity*, *commit*, *commercial*, and so on. Mis-stressed *borrow* could similarly activate initially unstressed words such as *barometer* or *baronial* or *bereft*. The intended word would not be among the listeners' cohort of initially activated lexical candidates. Moreover, English listeners' tendency to assume stressed syllables to be word-initial could result in temporary activation of word candidates beginning with the erroneously stressed second syllables *-mon* and *–row*, for example, *monitor* or *rowing*.

Analogous problems arise with a shift of stress in a word that, correctly spoken, would have a reduced vowel in the first syllable. Thus mis-stressed *October* would activate *octopus*, *octave*, *octane* (and for listeners from some dialect areas, such as the author's own Australian English, *auction*, *okra*, and *ocker* as well). Mis-stressed *addition* will activate *additive*, *addle*, *adder*, or *adamant*. Once again, in each case the initially activated set of candidate words contains a misleading array of words unrelated to what the speaker intended to say.

Finally, serious confusion will also arise even with an error in which the stress is correctly assigned but a reduced vowel is produced as a full vowel: *delay* in which the first syllable is compatible with that of *decent* or *dealer*, *number* in which the second syllable is compatible with the beginning of *burning* or *birthday*. Once again, the English listener's overlearned tendency to treat every full syllable as a potential word onset will result in two sets of lexical candidates where, with correct pronunciation, there should have been just one. Given the role that vowel reduction plays in stress realization, such mispronunciations are indeed errors of stress.

All such mis-stressings will, then, certainly delay recognition of the intended word. It may not rule it out; we do usually work out what people mean when they make a slip of the tongue, or when part of what they have said is inaudible. Indeed, mispronunciations of vowels are actually easier for listeners to recover from than mispronunciations of consonants (Cutler et al. 2000). This is because, in running speech, vowels are influenced by the consonants that abut them to a greater extent than consonants are influenced by adjacent vowels, and this asymmetry has led listeners to build up experience with having to alter initial decisions about vowels

more often than initial decisions about consonants. (The ability to adjust decisions about vowels is also, of course, handy in dealing with speakers from other dialectal areas, given that, in English, vowels are the principal carriers of dialectal variation. Not all speakers of English have the same vowel in the first syllable of *auction*, *okra*, and *octave*; see the previous section on the production of English lexical stress by native listeners for far more on this topic). Mis-stressing that includes mispronunciation of a vowel will activate an initial set of word candidates in which the intended word is not included, and further processing of the incoming speech will probably fail to produce a matching interpretation. The listener will have to reset the vowel interpretation and reanalyze; thus recognition will be delayed.

It is also significant that when native English-speakers make slips of the tongue that shift stress, the result will be most likely to activate a word that is very closely related to the intended word – *certificate* instead of *certification*, and so on. The effect will be to make accessible some aspects of the relevant meaning anyway and reanalysis is likely to be far swifter in such a case.

Lexical stress and non-native use of English

Both the production and perception of English lexical stress can offer problems, directly or indirectly, to the non-native user. In speech production, non-native users whose native phonology has no distinctions of stress face the challenge of pronouncing English stress in a native-like manner. In fact even learners whose native language has stress, but realizes it in a different way from the English, can be challenged by this task, whether the native language of the learner in question has fixed stress placement or has lexical stress that is realized purely suprasegmentally (see, for example, Archibald 1997; Guion, Harada, and Clark 2004; Peperkamp and Dupoux 2002). Indeed, even with both suprasegmental and segmental reflections of stress, two languages can differ in the relative strength of stress realization in each dimension, which can again complicate the acquisition of accurate pronunciation (Braun, Lemhöfer, and Mani 2011).

As the evidence summarized in the second section of this chapter makes clear, however, the most important production challenge that English lexical stress poses for the non-native user is actually a segmental one. English native listeners pay attention to whether vowels are full or reduced and use this information not only to identify words but also to segment running speech into its component words. The primary challenge therefore is not to utter a full vowel when the target utterance requires a reduced vowel, since this – as laid out in the previous section on mispronunciation of stress – is exactly what will mislead native listeners and potentially cause them to make inappropriate assumptions about where word boundaries are located. (Thus if the word *target* is uttered with correctly placed stress on the initial syllable, but with the second syllable unreduced – so that it sounds like *get* – it is liable to be perceived as two words rather than one; the same will happen if in correctly stressed *utterance* either its second or third syllable is not

reduced.) Non-native speakers of English from a variety of language backgrounds do indeed produce full vowels where reduced vowels would be called for (e.g., Fokes and Bond 1989; Zhang, Nissen, and Francis 2008). Native listeners' comprehension is then indeed affected by this. Braun, Lemhöfer, and Mani (2011) had British English and Dutch talkers produce English words such as *absurd*, *polite* (with an unstressed initial syllable), and used these in a word recognition task like those of Soto-Faraco, Sebastián-Gallés, and Cutler (2001) and others described in the second section above. Auditory presentation of the initial syllables (e.g., *ab-*) of native talkers' productions significantly assisted British English listeners' subsequent recognition of the matching complete words; the initial syllables from the Dutch talkers' productions (much less reduced than the native talkers' syllables) did not facilitate word recognition at all.

The stress production picture has another side, however, that is also shown by the evidence documented in the previous section; if a non-native user of English incorrectly assigns stress (without altering the pattern of full and reduced vowels), this may not even be noticed by native listeners, and in any case is unlikely to cause them comprehension problems. (Primary stress should fall on the first syllable of *SUMmarise* and on the third syllable of *inforMAtion*, but the evidence from the studies of mis-stressing suggests that listeners will also succeed in identifying *summaRISE* or *INformation*, with the correct vowels but misplaced primary stress location.)

In perception, non-native listeners will bring to speech input all the useful strategies that long experience with their native language has encouraged them to develop (Cutler 2012). These may or may not match the listening strategies encouraged by the probabilities of English; where they do not match, they will generate speech perception difficulty unless listeners can succeed in inhibiting their use. At the word recognition level, such perceptual problems fall into three principal groups: pseudo-homophony, spurious word activation, and temporary ambiguity.

Pseudo-homophones are words that are distinguished by some contrast that a non-native listener does not perceive: If English /r/ and /l/ cannot be distinguished, then *wrap* and *lap* become homophones. Pseudo-homophones are not a serious problem for the non-native listener (or indeed for native listeners processing non-native pronunciation), simply because, as discussed in the second section above, every language contains many homophones and all listeners have to be able to understand them by choosing the interpretation appropriate to the context. There is no way to understand the utterances *It's a mail* and *It's a male* except in relation to the discourse context. Given the extent of homophony in the English vocabulary, the number of homophones added by any one misperceived phonemic contrast is trivial (Cutler 2005b). Stress minimal pairs are especially rare; for a non-native listener who cannot hear a stress difference in *INsight* versus *inCITE*, these words will become homophones, but as we saw, they are effectively homophones for native listeners too (Cutler 1986; Small, Simon, and Goldberg 1988).

Spurious lexical activation and prolonged ambiguity are more serious problems. The first occurs when embedded "phantom words" are activated for the non-native listener and produce competition that native listeners are not troubled

by; remaining with the /r/-/l/ phonemic contrast, an example is competition from *leg* in *regular*. Such extra activation and competition has been abundantly demonstrated in non-native listening (Broersma 2012; Broersma and Cutler 2008, 2011). The second occurs when competition is resolved later for the non-native than for the native listener (e.g., *register* is distinguished from *legislate* only on the sixth phoneme, rather than on the first). This phenomenon has also been extensively documented (Cutler, Weber, and Otake 2006; Weber and Cutler 2004). Misperception of lexical stress by non-native users could in principle lead to such problems of competition increase, for example, if native expectations assume that stress placement is fixed and appropriate lexical candidates match to part of the input (thus while native listeners would segment *that's likely to boomerang on you* at the stressed *boo-*, expectation of final stress might lead to activation of *taboo* and *meringue*). Such issues have not yet been investigated with the empirical techniques for examining lexical competition.

In perception as in production, however, the literature again suggests that there is a second side to the non-native stress story. A non-native user whose first language encourages attention to suprasegmental cues to stress could apply the fruits of this language experience to English; even though English listeners do not use such cues, English speakers certainly provide them (Cutler et al. 2007). Indeed, in judging the stress level of excised or cross-spliced English syllables, native speakers of Dutch (whose language requires attention to suprasegmental stress cues) consistently outperform native English listeners (Cooper, Cutler, and Wales 2002; Cutler 2009; Cutler et al. 2007). Although the English vocabulary does not deliver sufficient lexical payoff for native listeners to exploit the suprasegmental cues to stress, it is conceivable that non-native listeners who are able to use them could thereby derive some compensation for the competition increases caused by other listening shortcomings.

Conclusion

In phonology, lexical stress in English is encoded to a significant extent in the segmental patterning of a word; it does not act principally to distinguish one word from another; but it does provide highly useful cues to listeners as to where word boundaries are to be located in speech signals. In speech production, pronunciation of English lexical stress is thus a multi-dimensional exercise: the segmental sequence is produced along with a code for a primary stress location, which is used in computing the metrical pattern of the utterance as a whole. In speech perception, listeners attend primarily to the segmental sequence in identifying words and use the rhythmic patterning of full and reduced vowels to segment speech.

For the non-native speaker of English, the pronunciation patterns described in this chapter, and their perceptual consequences, potentially present both good news and bad. The good news is that stress errors that are purely suprasegmental may be uttered with impunity, as English listeners hardly attend to suprasegmental patterning. The bad news is that any stress error resulting in a mispronounced

vowel – and most stress errors do have this effect – will throw the native listener into mis-segmentation and at least temporary lexical confusion.

Acknowledgments

Thanks to Janise Farrell for comments on an earlier draft of the text.

NOTES

1 There are some well-known examples of "stress shift" in English, which have been written about quite a lot simply because they are such eye-catching violations of what is otherwise a very strict rule of English phonology. In words with stress on the second syllable and full vowels in both the first and second syllables, such as *typhoon* or *thirteen*, the rhythm of the context can alter the apparent degree of salience of the stressed syllable, which is more salient in *typhoons are coming*! and *the number thirteen*, less salient in *Typhoon Thomas* or *thirteen hundred*, where the immediately following syllable is stressed. Acoustic analyses have shown that the name "stress shift" is actually unjustified (see, for example, Shattuck-Hufnagel, Ostendorf, and Ross 1994). It is also possible to apply contrastive stress to otherwise unstressed morphological components of words, especially prefixes, and especially for humorous effect: *This whisky wasn't EXported, it was DEported*!

2 Note that the stress-shifting cases described in Note 1 all tend to INcrease the frequency of initial stress, rather than DEcreasing it. A conspiracy favoring the majority pattern may be suspected.

REFERENCES

Archibald, J. 1997. The acquisition of English stress by speakers of non-accentual languages: lexical storage versus computation of stress. *Linguistics* 35: 167–181.

Bolinger, D. 1981. *Two Kinds of Vowels, Two Kinds of Rhythm*, Indiana University Linguistics Club.

Bond, Z.S. 1981. Listening to elliptic speech: pay attention to stressed vowels. *Journal of Phonetics* 9: 89–96.

Bond, Z.S. and Garnes, S. 1980. Misperceptions of fluent speech. In: *Perception and Production of Fluent Speech*, R.A. Cole (ed.), 115–132. Hillsdale, NJ: Erlbaum.

Bond, Z.S. and Small, L.H. 1983. Voicing, vowel and stress mispronunciations in continuous speech. *Perception and Psychophysics* 34: 470–474.

Braun, B., Lemhöfer, K., and Mani, N. 2011. Perceiving unstressed vowels in foreign-accented English. *Journal of the Acoustical Society of America* 129: 376–387.

Broersma, M. 2012. Increased lexical activation and reduced competition in second-language listening. *Language and Cognitive Processes* 27: 1205–1224.

Broersma, M. and Cutler, A. 2008. Phantom word recognition in L2. *System: An International Journal of Educational Technology and Applied Linguistics* 36: 22–34.

Broersma, M. and Cutler, A. 2011. Competition dynamics of second-language listening. *Quarterly Journal of Experimental Psychology* 64: 74–95.

Brouwer, S., Mitterer, H., and Huettig, F. 2012. Speech reductions change the dynamics of competition during spoken word recognition. *Language and Cognitive Processes* 27: 539–571.

Browman, C.P. 1978. Tip of the tongue and slip of the ear: implications for language processing. *UCLA Working Papers in Phonetics* 42: i–149.

Cole, R.A. and Jakimik, J. 1980. How are syllables used to recognize words? *Journal of the Acoustical Society of America* 67: 965–970.

Cole, R.A., Jakimik, J., and Cooper, W.E. 1978. Perceptibility of phonetic features in fluent speech. *Journal of the Acoustical Society of America* 64: 44–56.

Colombo, L. and Zevin, J.D. 2009. Stress priming in reading and the selective modulation of lexical and sub-lexical pathways. *PLoS ONE* 4: e7219.

Cooper, N., Cutler, A., and Wales, R. 2002. Constraints of lexical stress on lexical access in English: evidence from native and nonnative listeners. *Language and Speech* 45: 207–228.

Cummins, F. and Port, R.F. 1998. Rhythmic constraints on stress timing in English. *Journal of Phonetics* 26: 145–171.

Cutler, A. 1980a. Errors of stress and intonation. In: *Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen and Hand*, V.A. Fromkin (ed.), 67–80, New York: Academic Press.

Cutler, A. 1980b. Syllable omission errors and isochrony. In: *Temporal Variables in Speech*, H.W. Dechert and M. Raupach (eds.), 183–190, The Hague: Mouton.

Cutler, A. 1986. *Forbear* is a homophone: lexical prosody does not constrain lexical access. *Language and Speech* 29: 201–220.

Cutler, A. 2005a. Lexical stress. In: *The Handbook of Speech Perception*, D.B. Pisoni and R.E. Remez (eds.), 264–289, Oxford: Blackwell.

Cutler, A. 2005b. The lexical statistics of word recognition problems caused by L2 phonetic confusion. In: *Proceedings of INTERSPEECH 2005*, Lisbon, September, 413–416.

Cutler, A. 2009. Greater sensitivity to prosodic goodness in non-native than in native listeners. *Journal of the Acoustical Society of America* 125: 3522–3525.

Cutler, A. 2012. *Native Listening: Language Experience and the Recognition of Spoken Words*, Cambridge, MA: MIT Press.

Cutler, A. and Butterfield, S. 1992. Rhythmic cues to speech segmentation: evidence from juncture misperception. *Journal of Memory and Language* 31: 218–236.

Cutler, A. and Carter, D.M. 1987. The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language* 2: 133–142.

Cutler, A. and Clifton, C. 1984. The use of prosodic information in word recognition. In: *Attention and Performance X: Control of Language Processes*, H. Bouma and D.G. Bouwhuis (eds.), 183–196, Hillsdale, NJ: Erlbaum.

Cutler, A. and Norris, D. 1988. The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance* 14: 113–121.

Cutler, A., Norris, D., and Sebastián-Gallés, N. 2004. Phonemic repertoire and similarity within the vocabulary. In: *Proceedings of the 8th International Conference on Spoken Language Processing*, S.H. Kim and M.Jin Bae (eds.), vol. 1, 65–68, Seoul: Sunjin Printing Co. (CD-ROM).

Cutler, A. and Pasveer, D. 2006. Explaining cross-linguistic differences in effects of lexical stress on spoken-word recognition. In: *Proceedings of the 3rd International Conference on Speech Prosody*, R Hoffman and H. Mixdorff (eds.), 237–240, Dresden: TUDpress.

Cutler, A., Weber, A., and Otake, T. 2006. Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics* 34: 269–284.

Cutler, A., Sebastián-Gallés, N., Soler-Vilageliu, O., and Ooijen, B. van 2000. Constraints of vowels and consonants on lexical selection: cross-linguistic comparisons. *Memory and Cognition* 28: 746–755.

Cutler, A., Wales, R., Cooper, N., and Janssen, J. 2007. Dutch listeners' use of suprasegmental cues to English stress. In: *Proceedings of the Sixteenth International Congress of Phonetic Sciences*, Saarbrücken, 1913–1916.

Donselaar, W. van, Koster, M., and Cutler, A. 2005. Exploring the role of lexical stress in lexical recognition. *Quarterly Journal of Experimental Psychology* 58A: 251–273.

Fear, B.D., Cutler, A., and Butterfield, S. 1995. The strong/weak syllable distinction in English. *Journal of the Acoustical Society of America* 97: 1893–1904.

Fokes, J. and Bond, Z.S. 1989. The vowels of stressed and unstressed syllables in nonnative English. *Language Learning* 39: 341–373.

Fromkin, V.A. 1976. Putting the emPHAsis on the wrong sylLABle. In: *Studies in Stress and Accent*, L. Hyman (ed.), 15–26, Los Angeles, CA: University of Southern California.

Grainger, J., Van Kang, M.N., and Segui, J.(2001. Cross-modal repetition priming of heterographic homophones. *Memory and Cognition* 29: 53–61.

Guion, S.G., Harada, T., and Clark, J.J. 2004. Early and late Spanish–English bilinguals' acquisition of English word stress patterns. *Bilingualism: Language and Cognition* 7: 207–226.

Kim, J., Davis, C., and Cutler, A. 2008. Perceptual tests of rhythmic similarity: II. Syllable rhythm. *Language and Speech* 51: 342–358.

Kuijpers, C. and Donselaar, W. van. 1998. The influence of rhythmic context on schwa epenthesis and schwa deletion. *Language and Speech* 41: 87–108.

Ladefoged, P. 2006. *A Course in Phonetics*, Boston, MA: Thomson/Wadsworth.

Lehiste, I. 1970. *Suprasegmentals*, Cambridge, MA: MIT Press.

Lehiste, I. and Peterson, G. 1959. Vowel amplitude and phonemic stress in American English. *Journal of the Acoustical Society of America* 31: 428–435.

Levelt, W.J.M. 1989. *Speaking: From Intention to Articulation*, Cambridge, MA: MIT Press.

Levelt, W.J.M. 1992. Accessing words in speech production: Stages, processes and representations. *Cognition* 42: 1–22.

Levelt, W.J.M., Roelofs, A., and Meyer, A.S. 1999. A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22: 1–38.

Lieberman, P. 1963. Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech* 6: 172–187.

Liberman, M. and Prince, A. 1977. On stress and linguistic rhythm. *Linguistic Inquiry* 8: 249–336.

McQueen, J.M. 2007. Eight questions about spoken-word recognition. In: *The Oxford Handbook of Psycholinguistics*, M.G. Gaskell (ed.), 37–53, Oxford: Oxford University Press.

McQueen, J.M. and Huettig, F. 2012. Changing only the probability that spoken words will be distorted changes how they are recognized. *Journal of the Acoustical Society of America* 131: 509–517.

Mehler, J., Dommergues, J.-Y., Frauenfelder, U., and Segui, J. 1981. The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior* 20: 298–305.

Murty, L., Otake, T., and Cutler, A. 2007. Perceptual tests of rhythmic similarity: I. Mora rhythm. *Language and Speech* 50: 77–99.

Otake, T., Hatano, G., Cutler, A., and Mehler, J. 1993. Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language* 32: 358–378.

Peperkamp, S., and Dupoux, E. 2002. A typological study of stress "deafness". In: *Papers in Laboratory Phonology VII*, C. Gussenhoven and N.L. Warner (eds.), 203–240, Berlin: De Gruyter.

Piantadosi, S.T., Tily, H., and Gibson, E. 2012. The communicative function of ambiguity in language. *Cognition* 122: 280–291.

Roelofs, A. and Meyer, A.S. 1998. Metrical structure in planning the production of spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24: 922–939.

Shattuck-Hufnagel, S. 1992. The role of word structure in segmental serial ordering. *Cognition* 42: 213–259.

Shattuck-Hufnagel, S., Ostendorf, M., and Ross, K. 1994. Stress shift and early pitch accent placement in lexical items in American English. *Journal of Phonetics* 22: 357–388.

Shattuck-Hufnagel, S., and Turk, A.E. 1996. A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research* 25: 193–247.

Shaw, J.A. 2012. Metrical rhythm in speech planning: priming or predictability. In: *Proceedings of the 14th Australasian International Conference on Speech Science and Technology (SST)*, Sydney, Australia.

Slowiaczek, L.M. 1990. Effects of lexical stress in auditory word recognition. *Language and Speech* 33: 47–68.

Slowiaczek, L.M. 1991. Stress and context in auditory word recognition. *Journal of Psycholinguistic Research* 20: 465–481.

Small, L.H. and Squibb, K.D. 1989. Stressed vowel perception in word recognition. *Perceptual and Motor Skills* 68: 179–185.

Small, L.H., Simon, S.D., and Goldberg, J.S. 1988. Lexical stress and lexical access: homographs versus nonhomographs. *Perception and Psychophysics* 44: 272–280.

Soto-Faraco, S., Sebastián-Gallés, N., and Cutler, A. 2001. Segmental and suprasegmental mismatch in lexical access. *Journal of Memory and Language* 45: 412–432.

Vitevitch, M.S. and Luce, P.A. 1998. When words compete: levels of processing in perception of spoken words. *Psychological Science* 9: 325–329.

Vitevitch, M.S., Luce, P.A., Charles-Luce, J., and Kemmerer, D. 1997. Phonotactics and syllable stress: implications for the processing of spoken nonsense words. *Language and Speech* 40: 47–62.

Weber, A. and Cutler, A. 2004. Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language* 50: 1–25.

Zhang, Y., Nissen, S.L., and Francis, A.L. 2008. Acoustic characteristics of English lexical stress produced by native Mandarin speakers. *Journal of the Acoustical Society of America* 123: 4498–4513.

7 The Rhythmic Patterning of English(es): Implications for Pronunciation Teaching

EE-LING LOW

Early research

This chapter provides an extensive review of early and recent research on rhythm, rhythm indices, and the measurement of rhythm in relation to different varieties of world Englishes. Implications of recent research on rhythm for pronunciation teaching will be considered.

A summary of early research studies on speech rhythm has been provided in Low (2006) and in Low (2014). This section takes reference from both these works. Early research on speech rhythm tended to focus on exactly which speech unit regularly recurs such that isochrony (or equality in timing) occurs. Based on the concept of whether stresses or syllables recurred at regular intervals, Pike (1945) and Abercrombie (1967: 69) classified languages into either being stress-timed or syllable-timed. For stress-timed languages, it is the feet (comprising one stressed syllable up to but not including the next stressed syllable) that contribute to the overall perception of isochrony in timing. In the case of syllable-timed languages, it is the syllables that are believed to contribute to the perception of isochrony. However, the concept of pure or perfect isochrony became a moot point in the 1980s, with scholars proposing that isochrony should be described as a tendency rather than as an absolute. Dauer (1983) and Miller (1984), for example, suggest a continuum of rhythmic typology where languages can fall in between being stress-based at one end and syllable-based on the other (Grabe and Low 2002; Low 2006, 2014).

The earliest works classifying the rhythmic typology of the world's languages tended to forward the strict dichotomous view where languages were considered as either being stress- or syllable-timed. Abercrombie (1965: 67) believed that it is the way that chest or stress pulses recur that helps determine the rhythmic typology of a language, and for stress-timed languages it was the stress pulses that were

The Handbook of English Pronunciation, First Edition. Edited by Marnie Reed and John M. Levis.

isochronous while for syllable-timed languages chest pulses were isochronous. A third categorization of rhythm, known as mora-timing, was proposed by another group of scholars (e.g., Bloch 1942; Han 1962; Hoequist 1983a, 1983b; Ladefoged 1975). Japanese is the only language that scholars classified as being mora-timed. Because mora-timing does not apply to English, its specific details will not be discussed further except where it is relevant in other studies detailing rhythm across languages.

While early scholars proposed stress-, syllable-, and mora-timing as a means for classifying rhythmic typology of the world's languages, research also highlighted clear difficulties in adopting these categorical distinctions. For example, when interfoot (interstress) intervals were measured for stress-timed languages, researchers could not find evidence that their timing was roughly equal, that is, isochronous (Shen and Peterson 1962; Bolinger 1965; Faure, Hirst, and Chafcouloff 1980; Nakatani, O'Connor, and Aston 1981; Strangert 1985; Lehiste 1990). Yet others tried to find evidence that syllables were more nearly equal in timing for syllable-timed languages but failed (Delattre 1966; Pointon 1980; Borzone de Manrique and Signorini 1983).

Roach (1982) and Dauer (1983) measured the interstress intervals of different languages classified as stress- and syllable-timed. Roach's (1982) research set out to test the claims made by Abercrombie (1967) that syllables do not vary in length for syllable-timed languages and that interstress intervals ought not be equal in timing compared to stress-timed languages. Not only did Roach not find evidence to support these two claims but he found evidence that contradicted earlier claims because there was greater variability for syllable durations for syllable-timed languages compared to stress-timed ones and interstress intervals varied more in stress-timed languages compared to their syllable-timed counterparts. Roach's findings led him to suggest that evidence for the rhythmic categorization of languages cannot be sought by measuring timing units like syllables or interstress intervals in speech. Dauer (1983) conducted a cross-linguistic study of English, Thai, Italian, Greek, and Spanish. She found that interstress intervals were not more equal in languages classified as stress-timed, like English compared to Spanish, which has been classified to be syllable-timed. She therefore reached the same conclusion as Roach where she concluded that empirical support for rhythmic categorization cannot be found by measuring timing units found in speech. This led other scholars like Couper-Kuhlen (1990, 1993) to forward the view that isochrony is better understood as a perceptual rather than an acoustically measurable phenomenon. The experimental findings for mora-timed languages yielded mixed results. Port, Dalby, and O'Dell (1987) found some evidence that mora was nearly equal in timing in Japanese but others could not (Oyakawa 1971; Beckman 1982; Hoequist 1983a, 1983b).

Due to the experimental findings by early researchers where empirical evidence for rhythmic categorization was not related to timing units in speech, isochrony was then considered to be a tendency. This led to the terms stress-based, syllable-based, and mora-based languages in place of the earlier categorization of stress-, syllable-, and mora-timed (Dauer 1983, 1987; Laver 1994: 528–529). Grabe and Low (2002: 518) forwarded the proposal that "true isochrony is assumed

to be an underlying constraint" while it is the phonetic, phonological, syntactic, and lexical characteristics of a language that are likely to affect the isochrony of speech units found in any language. These characteristics form the basis for later research attempting to hunt for acoustic validation for the rhythmic classification of the world's languages as being stress-, syllable-, or mora-based.

Recent research

Early experimental studies on rhythm were unable to find support for isochrony by measuring timing intervals in speech. This led to the hypothesis proposed by researchers like Dauer (1983, 1987) and Dasher and Bolinger (1982) that rhythmic patterning is reliant on other linguistic properties of language such as their lexical, syntactic, phonological, and phonetic attributes. Dauer singled out three main influences on speech rhythm: syllable structure, the presence or absence of reduced vowels, and the stress patterning of different languages. She suggested that stress-based languages tend to have a more complex syllable structure make-up, and syllable-based languages also tend not to make a strong distinction between full and reduced vowels. Dasher and Bolinger (1982) also observed that syllable-based languages tended not to have phonemic vowel length distinctions, i.e., that long versus short vowels were not used as distinct phonemes, leading to long/short vowel conflations.

Nespor (1990) introduced the concept of "rhythmically mixed" or intermediate languages. For her, the strict categorical distinction was no longer tenable and languages were mainly mixed or intermediate in terms of rhythmic typology, and so-called intermediate languages exhibited shared properties characteristic of both stress- and syllable-based languages. One example of an intermediate language is Polish, which tends to be classified as being stress-based but which does not have reduced vowels, a feature that helps stress-based languages to achieve foot isochrony through compensatory shortening of syllables. Catalan is another such language, which has been classified as syllable-based but which has vowel reduction, a property that is not usually found in syllable-based languages.

Rhythm indices and the measurement of rhythm of world Englishes

The hunt for empirical acoustic validation for rhythmic classification of the world's languages led researchers to measure the durations of some phonological properties such as vowels, syllables, or consonants. In tandem with this focus on measuring durational units in speech, several rhythm indices have been developed to capture the rhythmic patterning of different languages, as indicated by the durational properties of the different timing units in speech. A nonexhaustive summary of the main rhythmic indices developed from the late 1990s to the present will be presented here. Tan and Low (2014) also present a version of this summary of latest developments on speech rhythm using rhythm indices.

The key breakthrough in the development of rhythmic indices to measure timing intervals can be traced back to the pairwise variability index found in Low (1994, 1998) but published in Low, Grabe, and Nolan (2000). At about the same time, the rhythmic indices developed by Ramus, Nespor, and Mehler (1999), Deterding (2001), Grabe and Low (2002), and Dellwo and Wagner (2003) were also developed. The main contribution of the rhythm indices to the study of speech rhythm was to show that it was possible to find empirical evidence to classify rhythm by measuring timing intervals in speech and subjecting them to calculations made possible by the rhythmic indices. Low (2006) details the development of the earlier rhythmic indices. Ramus et al.'s (1999) index and Low et al.'s (2000) index were applied to successive consonantal and vowel intervals respectively. The earlier indices are premised on the fact that stress-based languages tend to have a greater difference durationally between stressed and unstressed syllables and have a more complex syllable structure with more consonantal clusters in the onset and coda positions. This in turn influences the overall consonantal durations, making them longer. Nolan and Asu (2009) note that one advantage of Ramus et al.'s (1999) interval measures (IM) and Low et al.'s (2000) Pairwise Variability Index (PVI) is that a researcher is able to measure the timing intervals of languages not known to them because there is no need to consider the phonological make-up of syllables. Instead, as long as one is able to segment the speech signal into vowels and consonantal intervals, it is possible to apply both these indices to their measurements with little difficulty.

To elaborate on these two indices, Ramus et al.'s (1999) IM concentrated mainly on three timing intervals that are said to vary durationally across different languages. %V measures the proportion of vocalic intervals in speech (the segment between the vowel onsets and offsets); ΔV measures the standard deviation of the vocalic intervals, while ΔC measures the standard deviation of consonantal intervals (the segment of speech between vowel offsets and onsets excluding any pauses). These three IM were applied to languages classified as stress-based (Polish, Dutch, and English), mora-based (Japanese) and syllable-based (Catalan, Spanish, Italian, and French). Their results showed that the most reliable way to classify rhythmic patterning is to use ΔV and a combination of either ΔC or %V. The problem with using either ΔC or ΔV is that standard deviations are unable to capture the successive durational patterning of successive timing intervals, be they vowels or consonants, as pointed out by Low, Grabe, and Nolan (2000).

The rhythm indices developed by Low, Grabe, and Nolan (2000) are known as the PVI. It measures the durational variation that exists between successive vowels found in an utterance. The PVI is premised on the hypothesis that the main difference between stress-based and syllable-based languages is the lack of contrast between full and reduced vowels in syllable-based languages. This hypothesis is further premised on the assumption that stress-based languages need to have compensatory shortening for feet that contain a lot of syllables so that they can approach foot iscochrony, a central property of stress-based languages. Compensatory shortening is achieved via reduced vowels in unstressed syllables. Low (1998) and Low, Grabe, and Nolan (2000) considered the claim by Taylor (1981)

that it is the vowels, not the syllables, that determined the syllable-based nature of Singapore English. They compared the successive vowel durations found in British English (a stress-based language) with Singapore English (a syllable-based language). The PVI measures the mean absolute difference between successive vowels in an utterance. Absolute differences in durations between pairs of successive vowels are calculated and their means are taken (only positive values are considered by disregarding the negative sign when negative values occur). The mean difference is then calculated by dividing the difference between the successive vowel durations by the durational average of the two pairwise vowels so as to normalize for different speaking rates. To produce whole numbers, the values are multiplied by 100 and expressed as an index while the formula may be represented as

$$nPVI = 100 \times \left[\sum_{k=1}^{m-1} \left| \frac{d_k - d_{k+1}}{(d_k + d_{k+1})/2} \right| / (m-1) \right]$$

where m = number of vowel intervals in an utterance and d = duration of the kth vowel.

As the PVI measures the variation between successive vowels in an utterance, it is possible to surmise that an idealized stress-based language ought to have a high PVI while an idealized syllable-based language will have a low PVI. The highest possible PVI showing maximal variation between successive timing units is 100 while the lowest possible PVI showing no variation between successive timing units is 0. Low, Grabe, and Nolan (2000) discovered a signficant difference in PVI values between British English and Singapore English and concluded that the greater variation in the successive vowel durational units contributed to the perception of British English as stress-based and, consequently, the lack of variation in successive vowel durational units contributed to the perception of Singapore English as syllable-based. Applying Ramus, Nespor, and Mehler's (1999) IM %V to the data, they did not find that this was useful in reflecting the rhythmic patterning of both language varieties. However, if we consider %V to be the proxy for syllable structure make-up, we can then conclude that the difference between stress- and syllable-based languages cannot be captured adequately by considering differences in syllable structure make-up. The main breakthrough in Low, Grabe, and Nolan's (2000) work is that a measure for empirically capturing the difference between stress- and syllable-based languages could be found by measuring timing intervals in the speech signal, namely successive vowel durations.

Grabe and Low (2002) extended the investigation to 18 different languages and used both the normalized vocalic PVI values (nPVI) and the raw PVI scores for consonants (rPVI) for the investigation of prototypically stress-based languages (Dutch, German, and English), prototypically syllable-based languages (Spanish and French), and a prototypically mora-timed language (Japanese). The nPVI and rPVI were also applied to Polish and Catalan (classified by Nespor as being rhythmically mixed or intermediate) and three languages whose rhythmic

patterning has never been classified (Greek, Estonian, and Romanian). Grabe and Low (2002) found further evidence to show that prototypically stress-based languages like German, Dutch, and English had higher normalized vocalic variability (nPVI) while prototypically syllable-based languages like Spanish and French tended to have a lower normalized vocalic variability (nPVI). The alternation of full and reduced vowels in stress-based languages was more prominent than in syllable-based languages. Japanese had an nPVI reading that was closer to stress-based languages but high consonantal rPVI, which resembled that found for syllable-based languages. Catalan showed traits of being rhythmically mixed, as suggested by Nespor (1990), because it has a high normalized vocalic PVI characteristic of stress-based languages but it also had a high consonantal raw PVI normally associated with syllable-based languages. The raw PVI for consonantal intervals also showed the ability to tease out further differences between different languages like Polish and Estonian, which had similar vocalic nPVI values but different rPVI consonantal values.

More recent work (Dellwo and Wagner 2003) on developing rhythmic indices has emphasized the importance of normalizing rhythm indices against speaking rates across the entire utterance. Dellow's (2006) index, known as VarcoC, measures the standard deviation of consonantal intervals and divides the value by the mean consonantal duration in order to normalize for speech rates. VarcoC (the normalized version of ΔC) was found to be more robust than ΔC in capturing the difference between stress-based and syllable-based languages. However, Dellwo and Wagner (2003) found that normalizing for speech rate does not affect successive vocalic durations significantly and that it is therefore more important to control for speech rates when measuring consonantal intervals. White and Mattys (2007a, 2007b) devised a VarcoV in spite of Dellwo and Wagner's suggestion that vocalic intervals need not be normalized for speech rate and found that VarcoV was able to show the influence of one's L1 rhythm when VarcoC cannot.

More recent work on rhythm indices combines more than one index to the data. Loukina et al. (2009) found that combining two rhythm indices was more effective at classifying rhythmic differences between languages but that combining three indices did not yield better results. Studies combining different indices abound in the literature (Gibbon and Gut 2001; Gut et al. 2001; Dellwo and Wagner 2003; Asu and Nolan 2005; Lin and Wang 2005; Benton et al. 2007). Only those studies that further our understanding of the rhythmic patterning in different varieties of English spoken around the world will be highlighted.

Ferragne and Pellegrino (2004) found that the nPVI of successive vocalic intervals was a good way to automatically detect the difference in the dialects of English spoken in the British Isles but that consonantal intervals or the rPVI of consonants was not effective in detecting dialectal differences. Other studies have examined the influence of a speaker's L1 rhythm on their L2 rhythm by combining the rhythm indices. Lin and Wang (2005) used a combination of ΔC and %V and showed that L2 speakers of Canadian English were influenced rhythmically by their L1 Mandarin Chinese. Mok and Dellow (2008) applied the following indices on their data, ΔV, ΔC, ΔS, %V, VarcoV, VarcoC, VarcoS, rPVI-C, rPVI-S, nPVI-V, and

nPVI-S, where S refers to syllable durations, and found that L2 speakers of English were influenced rhythmically by their L1 Cantonese and Beijing Mandarin. Carter (2005) found that the rhythm of American English L2 Spanish bilingual speakers who had moved from Mexico to North Carolina was influenced by the L1 rhythm of Mexican Spanish. The PVI values obtained were intermediate between what one would expect for a stress-based language like English and a syllable-based language like Spanish. Whitworth's (2002) study on English and German bilinguals showed that bilingual children in these two stress-based languages produced the same PVI values for English and German as their parents' respective first languages. White and Mattys (2007a) used the following rhythm indices, ΔV, ΔC, %V, VarcoV, VarcoC, nPVI, and cPVI, to compare the rhythmic patterning of L1 and L2 speakers of English, Dutch, Spanish, and French. They found VarcoV to be the best discriminator between L1 and L2 speech rhythms, as a significant difference in VarcoV was found between the two groups of speakers.

Even more recent research on speech rhythm has argued for measuring other timing units such as foot and syllable durations and to consider the measurement of intensity in addition to merely timing durations (Ferragne 2008; Nolan and Asu 2009). These studies have also argued for considering the notion of rhythmic coexistence where a language can be both stress-based and syllable-based simultaneously. The proposal for considering foot and syllable durations can be challenging as it is harder than segmenting vowels and consonants in the speech signal. Furthermore, in typical stress-based languages, foot segmentation is a real issue if stressed and unstressed syllables are not significantly contrasted.

The application of rhythm indices to measuring different varieties of world Englishes has continued in recent years. Low (2010) applied the nPVI to British English, Chinese English (by a speaker of Beijing Mandarin), and Singapore English. Findings showed that while Singapore English differed significantly from British English (corroborating earlier studies), Chinese English rhythm did not differ significantly from either Singapore or British English. These findings provide support for the Kachruvian notion that Inner Circle varieties like British English provide the norms that Expanding Circle varieties like Chinese English veer towards. However, what is interesting is that at least in the rhythmic domain, Chinese English also veered towards Outer Circle norms, like Singapore English, which are supposed to be norm-developing varieties.

Mok (2011) measured the consonantal, vocalic, and syllabic intervals of Cantonese–English bilingual children and their age-matched monolingual counterparts. Results showed that at least in the syllabic domain, bilingual English speakers exhibited less variability than monolingual English speakers and this could signal a delay in the acquisition of L2 rhythm. She suggests that the lack of a strong contrast between stressed and unstressed syllables and the absence of reduced vowels in Cantonese may account for the delay. There is also evidence of syllabic simplification of Cantonese spoken by the Cantonese–English bilinguals, showing that bilingual speakers also show delay in language acquisition in both L1 and L2.

Payne et al. (2011) compared the speech of English, Spanish, and Catalan children aged 2, 4, and 6 with that of their mothers and found that they had more vocalic intervals but less durational variability. By age 6, interestingly, the children acquired similar vocalic interval patterning as their mothers but significantly different consonantal components.

Nakamura (2011) discovered that the ratio of stressed to unstressed syllables was lower for non-native compared to native speakers of English, showing that less contrast between stressed and unstressed syllables can be found in non-native English speech. Nokes and Hay (2012) applied the PVI to measure variability in the durational, intensity, and pitch of successive vowels of New Zealand English (NZE) speakers born between 1951 and 1988. The cross-generational study showed that younger speakers of NZE tended to show less of a distinction between stressed and unstressed vowels.

In recent years, Multicultural London English (MLE), spoken by different migrants in the inner city of London, has received much attention. Togersen and Szagay (2012) compared the rhythmic patterning of MLE speech compared to outer city counterparts and found that the MLE speakers had significantly lower PVI values compared to their outer London peers. The lower PVI values and more syllable-based rhythmic patterning is consistent with L2 varieties of English spoken around the world. Diez et al. (2008) found that the higher the proficiency of the L2 speaker, the more native-like their rhythmic patterning is likely to be.

Implications for pronunciation teaching

This section will discuss the relevant studies on speech rhythm that help inform pronunciation teaching and learning. What is clear from the detailed literature review of the research is that L2 rhythm is clearly influenced by L1 rhythm. Earlier research by Grabe, Post, and Watson (1999) suggested that the rhythm of a syllable-timed language like French is easier to acquire than that of a stress-timed language like English. Their evidence was found through comparing the PVI values of 4-year-old French and English children with their mothers. While French 4-year-olds had statistically similar PVI values compared with their mothers, English children clearly did not. More recent research by Payne et al. (2011) showed that by age 6, all children effectively acquired the rhythmic patterning of their mothers. The two studies taken together suggest that the syllable-timed advantage in the acquisition of rhythmic patterning levels out by the time children reach 6 years of age. This suggests that in order to capitalize on this advantage, exposure to the spoken language(s) that the child needs to learn should start from 4 years or earlier.

Another set of findings has implications for the early treatment and diagnosis of speech disorders. Peter and Stoel-Gammon (2003) looked at the rhythmic patterning of two children suspected of childhood apraxia compared to healthy controls. They found that singing a familiar song, imitating clapped rhythms, and repetitively tapping showed significant differences. This suggests that comparing

the rhythmic patterning of healthy and impaired children speaking a first language can be used as a diagnostic test for childhood speech apraxia.

The rhythmic patterning of native versus non-native varieties of English also showed significant differences. PVI values for Singapore English (Low, Grabe, and Nolan 2000), Nigerian English (Gut and Milde 2002), and Hispanic English (Carter 2005) were significantly lower compared to British English speakers, showing therefore a more syllable-based tendency for non-native varieties of English. The lower PVI values are, at least in part, due to the lack of a strong contrast between full and reduced vowels. However, for teachers of pronunciation, it is important to point out that the absence of reduced vowels may in fact help rather than hinder intelligibility (Janse, Nooteboom, and Quene 2003).

In terms of the development of English as an international language (EIL), Low (2010) showed that Chinese English had similar rhythmic patterning as British English (previously described as norm-providing) and Singapore English (previously described to be norm-developing). This led me to put forward the suggestion that the Kachruvian three circles model for world Englishes requires a re-thinking, at least in the rhythmic domain. No longer is the division of world Englishes into three concentric circles relevant when, in fact, Expanding Circle varieties may display similar attributes to both Inner and Outer Circle varieties. One suggestion is the Venn diagram found in Low (2010), where there is an intersection of Expanding Circle varieties with the two other circles. In other words, the Inner and Outer Circles should not be contained one within the other but represent separate ends of a continuum. There is therefore the pull of the Expanding Circle both towards and away from Inner Circle norms depending on what the speakers are trying to portray or achieve with their language use. This finding has many important suggestions for reshaping the way we think about norms for pronunciation.

First of all, in the EIL classroom, there is a need to consider both local and global norms. Upholding either a local or global norm has different implications. Alsagoff (2007: 39) uses Singapore English as an example to demonstrate the difference between a globalist or localist orientation in the use of a language variety. The global or international variety is associated with "socio-cultural capital, camaraderie, informality, closeness and community membership". In terms of the EIL pronunciation classroom and in considering instruction on speech rhythm in particular, if learners aspire towards a globalist orientation then stress-based timing should be taught. However, if learners aspire towards a localist orientation, then syllable-based timing should be the focus of the pronunciation classroom. The key here is to introduce the element of choice to the learners, allowing them to decide their identity and orientation in the EIL pronunciation classroom.

Moving to the pragmatic norms in EIL pronunciation instruction, Deterding (2012) cites Crystal's (1995) suggestion that syllable-based timing is sometimes used by British English speakers to express irritation or sarcasm. In the EIL pronunciation classroom, instructors do need to point out the pragmatic implications when native speakers shift from stress-based to syllable-based timing so as to avoid misunderstandings in cross-cultural speech settings involving high stakes, such as in educational or business settings.

Those who argue for the importance of teaching stress-based rhythm state that it is important to achieve fluency (Cruttenden 2008) since in native varieties of English, the presence or absence of reduced vowels forms the lowest level of the prosodic hierarchy (Beckman and Edwards 1994). This is a view that is echoed by Teschner and Whitley (2004), who state that the sound system of the English language is based on the alternation of strong and weak syllables or stressed and unstressed syllables. Celce-Murcia, Brinton, and Goodwin (1996) also emphasize that stress-based rhythm helps improve the fluency of the speech of learners of English. Wong (1987: 21) considers rhythm to be one of the major "organizing structures that native speakers rely on to process speech"; thus deviation from the native-like rhythm of English might potentially lead native speakers not to fully understand the speech of non-native speakers of English who use a primarily syllable-based timing.

In the EIL classroom, there is a need to introduce both the concepts of stress-based and syllable-based timing and to point out which varieties of English exhibit stress-based or syllable-based tendencies. This is because in the EIL paradigm, it is important to note who one wishes to be understood by, and in some cases stress-based timing is important for achieving intelligibility but in other speech situations syllable-based timing might be more important.

On a final note, the fact that there are more non-native speakers than native speakers of English in the world and that China alone has about 400 million speakers of English suggests that syllable-timed rhythm of Asian varieties may well become the target model for global trade given the rising economic dominance of the region. It is therefore important to emphasize to pronunciation instructors the multirhythmic models available and the need to take student needs and local and global constraints into account when teaching rhythm.

REFERENCES

Abercrombie, D. 1965. *Studies in Phonetics and Linguistics*, London: Oxford University Press.

Abercrombie, D. 1967. *Elements of General Phonetics*, Edinburgh: Edinburgh University Press.

Alsagoff, L. 2007. Singlish: negotiating culture, capital and identity. In: *Language, Capital, Culture: Critical Studies of Language in Education in Singapore*, V. Vaish, S. Gopinathan, and Yongbing Liu (eds.), 25–46, Rotterdam: Sense Publishers.

Asu, E.L. and Nolan, F. 2005. Estonian rhythm and pairwise variability index. In: *Proceedings of Fonetik, 2005*, 29–32, Gothenburg. Accessed June 3, 2013: http://www.ling.gu.se/(en)/konferenser/fonetik2005/papers/Asu_Nolan.pdf.

Beckman, M.E. 1982. Segment duration and the "mora" in Japanese. *Phonetica* 39: 113–135.

Beckman, M.E. and Edwards, J. 1994. Articulatory evidence for differentiating stress categories. In: *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III*, P.A. Keating (ed.), 7–33, Cambridge University Press.

Benton, M., Dockendorf, L., Wenhua, J., Yang, L., and Edmondson, J. 2007. The continuum of speech rhythm: computational testing of speech rhythm of large corpora from natural Chinese and English speech. In: *Proceedings of ICPhS*, vol. 16, 1269–1272, Saarbrücken, Germany. Accessed June 10, 2013: http://.icphs2007.de/conference/Papers/1591/.

Bloch, B. 1942. Studies in colloquial Japanese IV: phonemics. *Language* 26: 86–125.

Bolinger, D.L. 1965. *Forms of English: Accent, Morpheme, Order*, Cambridge, MA: Harvard University Press.

Borzone de Manrique, A.M. and Signorini, A. 1983. Segmental reduction in Spanish. *Journal of Phonetics* 11: 117–128.

Carter, P.M. 2005. Quantifying rhythmic differences between Spanish, English, and Hispanic English. In: *Theoretical and Experimental Approaches to Romance Linguistics: Selected Papers from the 34th Linguistic Symposium on Romance Languages (Current Issues in Linguistic Theory 272)*, R.S. Gess and E.J. Rubin (eds.), 63–75, Amsterdam and Philadelphia, PA: John Benjamins.

Celce-Murcia, M., Brinton, D.M., and Goodwin, J.M. 1996. *Teaching Pronunciation: A Reference for Teachers of English to Speakers of Other Languages*, Cambridge: Cambridge University Press.

Couper-Kuhlen, E. 1990. Discovering rhythm in conversational English: perceptual and acoustic approaches to the analysis of isochrony, KontRI Working Paper 13.

Couper-Kuhlen, E. 1993. *English Speech Rhythm: Form and Function in Everyday Verbal Interaction*, Amsterdam: John Benjamins B.V.

Cruttenden, A. 2008. *Gimson's Pronuncitation of English*, 7th edition, London: Arnold.

Crystal, D. 1995. Documenting rhythmical change. In: *Studies in General and English Phonetics: Essays in Honour of Professor J.D. O'Connor*, J.W. Lewis (ed.), 174–179, London: Routledge.

Dasher, R. and Bolinger, D. 1982. On pre-accentual lengthening. *Journal of the International Phonetic Association* 12: 58–69.

Dauer, R.M. 1983. Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics* 11: 51–69.

Dauer, R.M. 1987. Phonetic and phonological components of language rhythm. In: *Proceedings of the 11th International Congress of Phonetic Sciences*, 447–450.

Delattre, P. 1966. A comparison of syllable length conditioning among languages. *International Review of Applied Linguistics in Language Teaching* 4(3): 183–198.

Dellwo, V. 2006. Rhythm and speech rate: a variation coefficient for DeltaC. In: *Language and Language Processing: Proceedings of the 38th Linguistic Colloquium*, P. Karnowski and I. Szigeti (eds.), 231–241, Piliscsaba 2003, Frankfurt: Peter Lang.

Dellwo, V. and Wagner, P. 2003. Relations between language rhythm and speech rate. In: *Proceedings of the 15th International Congress of the Phonetic Sciences*, M.J. Solé, D. Recasens, and J. Romero (eds.), 471–474, Rundle Mall: Causal Productions.

Deterding, D. 1994. The intonation of Singapore English. *Journal of the International Phonetic Association* 24(2): 61–72.

Deterding, D. 2001. The measurement of rhythm: a comparison of Singapore English and British English. *Journal of Phonetics* 29: 217–230.

Deterding, D. 2012. Issues in the acoustic measurement of rhythm. In: *Pragmatics and Prosody in English Language Teaching*, J. Romero-Trillo (ed.), 9–24, Dordrecht: Springer.

Diez, F.G., Dellwo, V., Gevalda, N., and Rosen, S. 2008. The development of measurable speech rhythm during second language acquisition. *Journal of the Acoustical Society of America* 123: 3886.

Faure, G., Hirst, D., and Chafcouloff, M. 1980. Rhythm in English: isochronism, pitch and perceived stress. In: *The Melody of Language*, L.R. Waugh and C.H. van Schooneveld (eds.), 71–79, Baltimore, MD: University Park Press.

Ferragne, E. 2008. Étude Phonétique des Dialectes Modernes de L'anglais des I1es Britanniques: Vers L'identification Automatique du Dialect. PhD Dissertation, Université Lyon.

Ferragne, E. and Pellegrino, F. 2004. A comparative account of the suprasegmental and rhythmic features of British English dialects. In: *Proceedings of "Modelisations pour I'Identification des Langues"*, Paris. Accessed July 2, 2013: http://www.limsi.fr/MIDL/actes/session%20IV/Ferragne&Pellegrino_MIDL2004.pdf.

Gilbert, J. 1978. Gadgets: nonverbal tools for teaching pronunciation. *CATESOL Occasional Papers* 4: 68–78.

Gibbon, D. and Gut, U. 2001. Measuring speech rhythm. In: *Proceedings of Eurospeech*, 91–94, Aalborg.

Grabe, E. and Low, E.-L. 2002. Durational variability in speech and the rhythm class hypothesis. In: *Laboratory Phonology 7*, C. Gussenhoven and N. Warner (eds.), 515–546, Berlin: Mouton de Gruyter.

Grabe, E., Post, B., and Watson, I. 1999. The acquisition of rhythmic patterns in English and French. In: *Proceedings of the International Congress of Phonetic Sciences*, 1201–1204.

Gut, U. and Milde J.-T. 2002. The prosody of Nigerian English. In: *Speech Prosody 2002*, 367–370.

Gut, U., Urua, E.-A., Adouakou, S., and Gibbon, D. 2001. Rhythm in West African tone languages: a study of Ibibio, Anyi and Ega. In: *Proceedings of Typology of African Prosodic Systems Workshop*, Bielefeld University, Germany, May 18–20.

Han, M.S. 1962. The feature of duration in Japanese. *Onsei no kenkyuu* 10: 65–80.

Hoequist Jr., C. 1983a. Durational correlates of linguistic rhythm categories. *Phonetica* 40: 19–31.

Hoequist Jr., C. 1983b. Syllable duration in stress-, syllable- and mora-timed languages. *Phonetica* 40: 203–237.

Janse, E., Nooteboom, S., and Quené, H. 2003. Word-level intelligibility of time-compressed speech: prosodic and segmental factors. *Speech Communication* 41: 287–301. doi:10.1016/S0167-6393(02)00130-9.

Kachru, B.B. 1992. Teaching world Englishes. In: *The Other Tongue: English across Cultures*, 2nd edition, B.B. Kachru (ed.), 355–365, Urbana, IL: University of Illinois Press.

Ladefoged, P. 1975. *A Course in Phonetics*, New York: Harcourt Brace Jovanovich.

Laver, J. 1994. *Principles of Phonetics*, Cambridge: Cambridge University Press.

Lehiste, I. 1990. Some aspects of the phonetics of metrics. In: *Nordic Prosody V*, K. Wiik and I. Ramo (eds.), 206–218, Turku: University of Turku.

Lin, H. and Wang, Q. 2005. Vowel quality and consonant variance: a comparison between Chinese and English. In: *Proceedings of between Stress and Tone*, Leiden.

Loukina, A., Kochanski, G., Shih, C., Keane, E., and Watson, I. 2009. Rhythm measures with language-dependent segmentation. In: *Proceedings of Interspeech 2009: Speech and Intelligence*, 1531–1534, Brighton, UK, 6–10 September, International Speech Communications Association.

Low, E.-L. 1994. Intonation Patterns in Singapore English. Masters Dissertation, University of Cambridge.

Low, E.-L. 1998. Prosodic Prominence in Singapore English. PhD Dissertation, University of Cambridge.

Low, E.-L. 2006. A review of recent research on speech rhythm: some insights for language acquisition, language disorders and language teaching. In: *Spoken English, TESOL and Applied Linguistics: Challenges for Theory and Practice*, R. Hughes (ed.), 99–125, Houndmills, Hampshire: Palgrave Macmillan.

Low, E.-L. 2010. The acoustic reality of the Kachruvian circles: a rhythmic perspective. *World Englishes* 29(3): 394–405.

Low, E.-L. 2014. *Pronunciation for English as an International Language: From Research to Practice*, London: Routledge.

Low, E.-L. and Grabe, E. 1999. A contrastive study of prosody and lexical stress placement in Singapore English and British English. *Language and Speech* 42(1): 39–56.

Low, E.-L., Grabe, E., and Nolan, F. 2000. Quantitative characterizations of speech rhythm: syllable-timing in Singapore English. *Language and Speech* 43(4): 377–401.

Miller, M. 1984. On the perception of rhythm. *Journal of Phonetics* 12: 75–83.

Mok, P.P.K. 2011. The acquisition of speech rhythm by three-year-old bilingual and monolingual children: Cantonese and English. *Bilingualism: Language and Cognition* 14: 458–472.

Mok, P.P.K. and Dellwo, V. 2008. Comparing native and non-native speech rhythm using acoustic rhythmic measures: Cantonese, Beijing Mandarin and English. In: *Speech Prosody 2008*, 423–426, Campinas, Brazil.

Nakamura, S. 2011. Characteristics of contrast between the stressed and the unstressed in rhythm units observed in duration structure in English speech by Japanese learners. *Pan-Pacific Association of Applied Linguistics* 15(1): 177–189.

Nakatani, L.H., O'Connor, J.D., and Aston, C.H. 1981. Prosodic aspects of American English speech rhythm. *Phonetica* 38: 84–105.

Nespor, M. 1990. On the rhythm parameter in phonology. In: *Logical Issues in Language Acquisition*, I. Roca (ed.), 157–175, Dordrecht: Foris.

Nokes, J. and Hay, J. 2012. Acoustic correlates of rhythm in New Zealand English: a diachronic study. *Language Variation and Change* 24: 1–31.

Nolan, F. and Asu, E.L. 2009. The Pairwise Variability Index and coexisting rhythms in language. *Phonetica* 66: 64–77.

Oyakawa, T. 1971. On the directionality of segmental conditioning in Japanese. *Monthly Internal Memo*, 81–103, University of Berkely, Califonia.

Payne, E., Post, B., Astruc, L., Prieto, P., and del Mar Vanrell, M. 2011. Measuring child rhythm. *Language and Speech* 55(2): 203–229.

Peter, B. and Stoel-Gammon, C. 2003. Rhythm production during speech and music tasks in two children with suspected childhood apraxia of speech (sCAS) and two typically developing children. Poster presentation at the Child Phonology Conference, Vancouver, BC, July 2–4, 2003.

Pike, K.L. 1945. *The Intonation of American English*, Ann Arbor, MI: University of Michigan Press.

Pointon, G.E. 1980. Is Spanish really syllable-timed? *Journal of Phonetics* 8: 293–304.

Port, R.F., Al-Ani, S., and Maeda, S. 1980. Temporal compensation and universal phonetics. *Phonetica* 37: 235–252.

Port, R.F., Dalby, F., and O'Dell, M. 1987. Evidence for mora-timing in Japanese. *Journal of the Acoustical Society of America* 81: 1574–1585.

Ramus, F., Nespor, M., and Mehler. J. 1999. Correlates of linguistic rhythm in the speech signal. *Cognition* 73(3): 265–292.

Roach, P. 1982. On the distinction between "tress-timed" and "syllable-timed" languages. In: *Linguistic Controversies*, D. Crystal (ed.), 73–79, London: Edward Arnold.

Shen, Y. and Peterson, G.G. 1962. Isochrony in English. *University of Buffalo Studies in Linsguistics Occasional Papers* 9: 1–36.

Strangert, E. 1985. *Swedish Speech Rhythm in a Cross-Linguistic Perspective*, Umea: *Umea Studies in Humanities 9*.

Tan, R.S.K., and Low, E.-L. 2014. Rhythmic patterning in Malaysian and Singapore English. *Language and Speech* 57(2): 196–214.

Taylor, D.S. 1981. Non-native speakers and the rhythm of English. *International Review of Applied Linguistics in Language Teaching* 19: 219–226.

Teschner, R.V. and Whitley, M.S. 2004. *Pronouncing English: A Stress-based Approach* (with CD-ROM), Washington, DC: Georgetown University Press.

Thomas, E.R. and Carter, P.M. 2006. Prosodic rhythm and African American English. *English World-Wide* 27: 331–355.

Torgersen, E.N., and Szakay, A. 2012. An investigation of speech rhythm in London English. *Lingua* 122: 822–840.

White, L. and Mattys, S.L. 2007a. Calibrating rhythm: first and second language studies. *Journal of Phonetics* 35: 501–522.

White, L. and Mattys, S.L. 2007b. Rhythmic typology and variation in first and second languages. In: *Segmental and Prosodic Issues in Romance Phonology (Current Issues in Linguistic Theory Series)*, P. Prieto, J. Mascaró, and M.-J. Solé (eds.), 237–257, Amsterdam, Philadelphia, PA: John Benjamins.

Whitworth, N. 2002. Speech rhythm production in three German–English bilingual families. In: *Leeds Working Papers in Linguistics and Phonetics* 9, D. Nelson (ed.), 175–205.

Wong, R. 1987. *Teaching Pronunciation: Focus on English Rhythm and Intonation*, Englewood Cliffs, NJ: Prentice-Hall Regents.

8 English Intonation – Form and Meaning

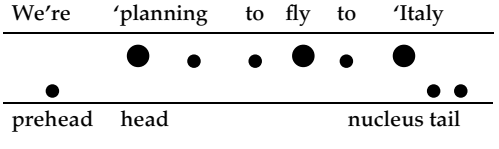
JOHN M. LEVIS AND ANNE WICHMANN

Introduction

Intonation is the use of pitch variations in the voice to communicate phrasing and discourse meaning in varied linguistic environments. Examples of languages that use pitch in this way include English, German, Turkish, and Arabic. The role of pitch in intonation languages is to be distinguished from its role in tone languages, where voice pitch also distinguishes meaning at the word level. Such languages include Chinese, Vietnamese, Burmese, and most Bantu languages. The aims of this chapter are twofold: firstly, to present the various approaches to the description and annotation of intonation and, secondly, to give an account of its contribution to meaning.

Descriptive traditions

There have been many different attempts to capture speech melody in notation: the earliest approach, in the eighteenth century, used notes on staves, a system already used for musical notation (Steele 1775 – see Williams 1996). While bar lines usefully indicate the phrasing conveyed by intonation, the stave system is otherwise unsuitable, not least because voice pitch does not correspond to fixed note values. Other less cumbersome notation systems have used wavy lines, crazy type, dashes, or dots to represent pitch in relation to the spoken text. See, for example, the representation in Wells (2006: 9) below, which uses large dots for accented syllables.



The Handbook of English Pronunciation, First Edition. Edited by Marnie Reed and John M. Levis.

An important feature of intonation is that not all elements of the melody are equally significant; the pitch associated with accented syllables is generally more important than that associated with unstressed syllables. This distinction is captured in the British system of analysis, which is built around a structure of phrases ("tone groups") that contain at least one accented syllable carrying pitch movement. If there are more than one, the last is known as the "nucleus" and the associated pitch movement is known as the "nuclear tone". These nuclear tones are described holistically as falling, rising, falling-rising, rising-falling, and level (Halliday proposed a slightly different set, but these are rarely used nowadays), and the contour also extends across any subsequent unstressed syllables, known as the tail. A phrase, or tone group, may also contain additional stressed syllables, the first of which is referred to as the onset. The stretch from the onset to the nucleus is the "head" and any preceding unstressed syllables are prehead syllables. This gives a tone group structure as follows: [prehead head nucleus tail] in which only the nucleus is obligatory. The structure is exemplified in the illustration above.

In the British tradition, nuclear tones are conceived as contours, sometimes represented iconically in simple key strokes [fall \, rise /, fall-rise \/, rise-fall/\, level -] inserted before the syllable on which the contour begins. This is a useful shorthand as in the following: *I'd like to \thank you | for such a \/wonderful ex\perience |*. In American approaches, on the other hand, pitch contours have generally been decomposed into distinct levels or targets, and the resulting pitch contour is seen as the interpolation between these points. In other words, a falling contour is the pitch movement between a high target and a low(er) target. These traditions, especially in language teaching, have been heavily influenced by Kenneth Pike (1945), whose system described intonation as having four pitch levels. Each syllable is spoken at a particular pitch level and the pattern of pitches identified the type of intonation contour. The primary contour (the British "nucleus" or the American "final pitch accent") is marked with ° . The highest possible pitch level is 1 and the lowest is 4. In the illustration below, Pike analyzed a possible sentence in two ways.

I want to go home	
3–	2°–4
I want to go home	
2–	2°–4

Both of these sentences are accented on the word *home* and both fall in pitch. Pike describes them as having different meanings, with the second (starting at the same level as *home*) as portraying "a much more insistent attitude than the first" (Pike 1945: 30). Intonational meaning, to Pike, was tightly bound up with communicating attitudes, and because there were many attitudes, so there had to be many intonational contours. A system with four pitch levels provided a rich enough system to describe the meanings thought to be communicated by intonation. Later researchers showed that Pike's system, ingenious as it was, overrepresented the number of possible contours. For example, Pike described many contours that

were falling and argued that they were all meaningfully distinct. The 3-2°-4 differed from the 2-2°-4, 1-2°-4, 3-3°-4, 2-3°-4, 3-1°-4, etc., although there is little evidence that English has so many falling contours with distinct meanings – the differences are more likely to be gradient ones expressing different degrees of affect. The system begun by Pike is used widely in American pronunciation teaching materials, including in the influential textbook *Teaching Pronunciation* (Celce-Murcia, Brinton, and Goodwin 2010).

Like the American tradition based on Pike's four pitch levels, the British system of nuclear tones has played a role in research (Pickering 2001) but more widely in teaching (e.g., O'Connor and Arnold 1973; Bradford 1988; Brazil 1994; Wells 2006).

The American notion of treating contours not holistically, as in the British tradition, but as a sequence of levels or targets, was developed further by Janet Pierrehumbert (1980), and this system forms the basis of the kind of analysis that is now most widely used in intonation research. It posits only two abstract levels, High (H) and Low (L). If the target is associated with a prominent (accented) syllable, the "pitch accent" is additionally marked with an asterisk [*], thus giving H* or L*. A falling nuclear tone would therefore be represented as H* L, in other words the interpolation between a high pitch accent (H*) and a low target (L), and a falling-rising tone would be represented as H* L H. Additional diacritics indicate a target that is at the end of a nonfinal phrase or the end of the sentence before a strong break in speech: intermediate phrases and intonational phrases.

This kind of analysis, referred to as the Autosegmental Metrical approach (see Ladd 1996) is now the norm in intonation research. Much of this research has been driven by the needs of speech technology, where a binary system (H and L) lends itself more readily to computer programming than any holistic analysis such as the British nuclear tone system. In addition, it leads to an annotation system that is easy to use in conjunction with instrumental analysis. However, while this combination of autosegmental phonology and acoustic analysis is common in speech research, it is not as common among applied linguists, where earlier American or British systems, e.g., Pike's four pitch levels, the British system of nuclear tones, together with auditory (impressionistic) analysis, remain the norm. This may be because these systems have a longer history, it may be because of their usefulness in language teaching, or it may be because applied researchers do not have familiarity with the research into intonation being carried out by theoretical and laboratory phonologists. The number of applied linguistic studies that have appealed to newer models of intonation is quite limited. Researchers such as Wennerstrom (1994, 1998, 2001) and Wichmann (2000) have provided accessible accounts of the pitch accent model for the applied researcher, but their work has, by and large, not been widely emulated in research and not at all in language teaching.

It is unlikely that intonation studies will ever dispense entirely with auditory analysis, but the greatest advance in the study of intonation (after the invention of the tape recorder!) has come with the widespread availability of instrumental techniques to complement listening. The advent of freely available speech analysis

software has revolutionized the field: published studies that do not make use of instrumental analysis are increasingly rare, and the ability to read and understand fundamental frequency contours in relation to waveform displays and sometimes spectrographic detail is an essential skill.

Instrumental analysis

The acoustic analysis of intonation involves the use of speech processing software that visualizes elements of the speech signal. There are three main displays that are useful in the study of intonation: the waveform spectrograms, and in particular the fundamental frequency (F0) trace, which is what we hear as speech melody or pitch. Interpreting the output of speech software requires some understanding of acoustic phonetics and the kind of processing errors that may occur. Figure 8.1 shows three displays combined – F0 contour, spectrogram, and waveform. It also shows the fragmentary nature of what we "hear" as a continuous melody. This is due to the segmental make-up of speech: only sonorant segments carry pitch (vowels, nasals, liquids) while fricatives and plosives leave little trace.[1] It is also common to see what seems like a sudden spike in pitch or short sequence much higher, or lower, than the surrounding contour. These are generally not audible to the listener. These are so-called "octave leaps", which are software-induced errors in calculating the F0 and are sometimes caused by the noise of fricatives or plosives. These examples show that it takes some understanding of the acoustic characteristics of individual speech sounds to read F0 contours successfully.

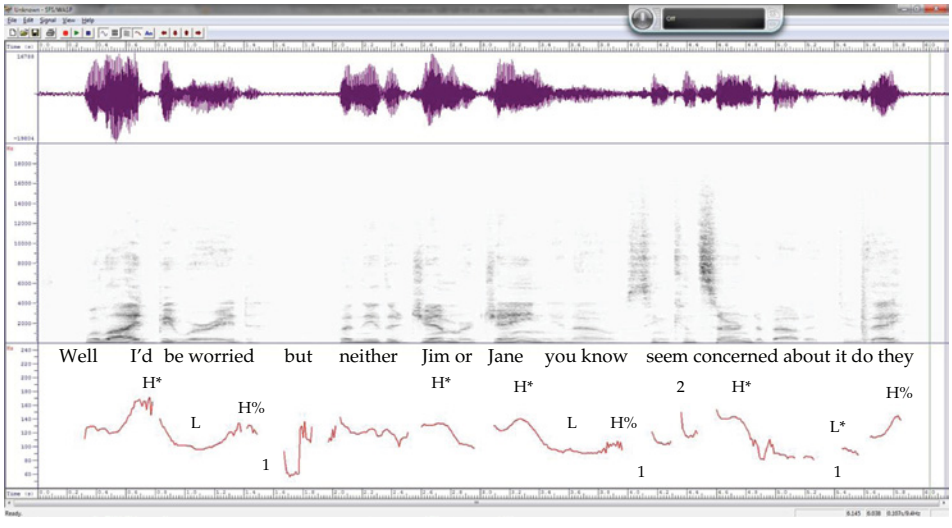


Figure 8.1 Well, I'd be worried / but neither Jim or Jane / you know / seem concerned about it / do they?

An important advance in the development of speech analysis software has been to allow annotation of the acoustic display so that the annotation and the display (e.g., spectrogram, waveform, F0 contour) are time-aligned. The most commonly used software of this kind, despite being somewhat user-unfriendly for anyone who does not use it on a daily basis, is Praat (Boersma 2001). This allows for several layers of annotation, determined by the user. Usually the annotation is in terms of auditorily perceived phonological categories (falls, rises, or more often H*L, L*H, etc.), but separate tiers can be used for segmental annotation or for nonverbal and paralinguistic features (coughs, laughter, etc.).

Working with the output of acoustic analysis of intonation is clearly not straightforward: we need to understand on the one hand something of acoustic phonetics and what the software can do, but on the other hand we need to realize that what we understand as spoken language is as much the product of our brains and what we know about our language as of the sound waves that we hear. This means that computer software can show the nature of the sounds that we perceive, but it cannot show us what we make of it linguistically.

The linguistic uses of intonation in English, together with other prosodic features, include:

1. Helping to indicate phrasing (i.e., the boundaries between phrases);
2. Marking prominence;[2]
3. Indicating the relationship between successive phrases by the choice of pitch contour (fall, rise, etc., or in AM terms the sequence of pitch accents). A phrase final fall can indicate finality or closure, while a high target such as the end of a rise or a fall-rise, can suggest nonfinality.

Phrases that make up an overall utterance are sometimes called tone units or tone groups. These correspond to a feature of the English intonation system called tonality (Halliday 1967). In language teaching, tone groups are often given other names as well, including thought groups or idea units, although such meaning-defined labels are not always helpful because it is not clear what constitutes an "idea" or a "thought". Each tone unit contains certain points in the pitch contour that are noticeably higher than others. These are syllables that will be heard as stressed or accented in the tone group. In English, these have a special role. In Halliday (1967), these are called tonic syllables and their system is called tonicity. The contour associated with each phrase-final accent ends with either high or low. These are examples of tonality in Halliday's system.

These three elements – phrasing, prominence placement, and contour choice – are part of intonational phonology. The H and L pitch accents are abstract – the phonology does not generally specify *how* high or *how* low, simply High or Low (or at best, higher or lower than what came before). However, the range of pitch over individual syllables, words, or longer phrases can be compressed or expanded to create different kinds of meaning. An expanded range on a high pitch accent can create added emphasis (*it's mine* versus *it's MINE!*), for example, or it can indicate a new beginning, such as a new paragraph or topic shift. A compressed

pitch range over a stretch of speech, on the other hand, may signal parenthetical information.

The display in Figure 8.1[3] illustrates features of English intonation with the acoustic measurement of an English sentence of 21 syllables. The top of the display shows the waveform, the middle the spectrographic display, and the bottom the fundamental frequency (F0) or pitch display. We will discuss a number of features visible in this figure.

Phrasing – boundaries and internal declination

Firstly, the sentence has four divisions as seen by the breaks in the pitch lines (marked with the number 1). These are, in this case, good indications of the way in which the sentence was phrased. It is important to note, however, that not all phrases are separated by a pause. In many cases the analyst has to look for other subtle signals, including pitch discontinuity and changes of loudness and tempo ("final lengthening") and increased vocal fry ("creak") to find acoustic evidence of a perceived boundary.

There is a second element of intonation present in this sentence, and that is the tendency of voice pitch to start high in a tone group and move lower as the speaker moves through the tone unit. This is known as declination and is clearest in the second tone group, which starts relatively high and ends relatively low in pitch. Related to this is the noticeable reset in pitch at the beginning of the next tone group (marked with 2). The only phrase to reverse this is the final phrase, which is a tag question. Tag questions can be realized with a fall or a rise, depending on their function. The rising contour here suggests that it is closer to a real question than simply a request for confirmation.

Prominence

The next linguistic use of pitch in English is the marking of certain syllables as prominent. In the AM system, these prominent syllables are marked as having starred pitch accents. Pitch accents (or peak accent; see Grice 2006) are, at their most basic, marked with either High pitch (H*) or Low pitch (L*).

This corresponds to other terminology including (in the British system): tonic (Halliday 1967), nuclear stress (Jenkins 2000), sentence stress (Schmerling 1976), primary phrase stress (Hahn 2004), focus (Levis and Grant 2003), prominence (Celce-Murcia, Brinton, and Goodwin (2010), highlighting (Kenworthy 1987), and selection (Brazil 1995), among others. The perception of prominence is triggered primarily (in English) by a pitch excursion, upwards or occasionally downwards. Again, pitch works together with other phonetic features in English to signal prominence, especially syllable lengthening and fuller articulation of individual segmentals (vowels and consonants), but pitch plays a central role in marking these syllables. The pitch excursions are often visible in the F0 contour – in Figure 8.1 they are aligned with the accented syllables; *I'd*, *Jim*, *Jane* and the second syllable of *concerned* all have H* pitch accents and are marked with′ while *do*

has a L* pitch accent – the beginning of a rising contour that reflects the questioning function of *do they*.

In the AM system, all pitch accents are of equal status, but the British system of nuclear tones reserves special significance for the last pitch accent in a phrase (or tone group). In the sentence in Figure 8.1, the second tone unit has two pitch accents, *Jim* and *Jane*. *Jane* would carry the nuclear accent, while the accent on *Jim* would be considered part of the Head.

The nuclear syllable is associated with the nuclear tone or pitch contour, described in holistic terms as a fall, a rise, or fall-rise, for example. The tone extends across any subsequent unstressed syllables up to the end of the tone group. In Figure 8.1 the nuclear fall beginning on *I'd* extends over *be worried.* This nuclear tone drops from the H* to an L pitch and then rises to the end of the tone group, with a final H% (the % means the final pitch level of a tone group). This is the kind of intonation that, when spoken phrase finally, "has a 'but' about it" (Cruttenden 1997; Halliday 1967: 141). That beginning on *concerned* extends over *about it*. Less easy to determine is the contour beginning on *Jane*: it could be a falling tone that flattens out over *you know* or, if the slight rise visible at the end of the phrase is audible, it could be another falling-rising contour (H*LH%, as marked) or a falling contour (H*LL%) extending across the three syllables *Jane you know*. Alternatively, the tone group can be seen as two tone groups, the first, *but neither Jim or Jane* followed by a separate phrase *you know*, particularly if the final item such as a discourse marker is separated by a slight pause – not the case here. In this analysis, *you know* would be an anomalous phrase with no nucleus and spoken at a low, level pitch with a slight rise or with a fairly flat contour. This kind of parenthetical intonation patttern was discussed by Bing (1980) and others (e.g., Dehé and Kavalova, 2007).

English also has other pitch accents that are characterized by the way that the prominent syllable aligns with the pitch accent. In the currently most fully developed system for transcribing intonation, the ToBI system, based on the work of Pierrehumbert (1980) and Beckman and Pierrhumbert (1986), English pitch accents can also be described as L+H*, L*+H, and H+!H*. These somewhat intimidating diacritics simply mean that the pitch accent is not perfectly aligned with the stressed syllable that is accented. In H* and L*, the vowel that is accented is aligned with the peak or lowest point of the pitch accent. This misalignment of pitch accent and stressed syllables is linguistically meaningful in English (see Ladd 2008; Pierrehumbert 1980; Pierrehumbert and Hirschberg 1992 for more information). Figure 8.2 shows the difference between the H* and L*+H pitch accent, which starts low on the word *I* but continues to a high pitch on the same syllable.

Discourse meaning

The sentence in Figure 8.1 has other features that are important in understanding English intonation. The first pitch accent, on *I'd*, is noticeably higher than the other pitch accents. This may be because it is first and because pitch

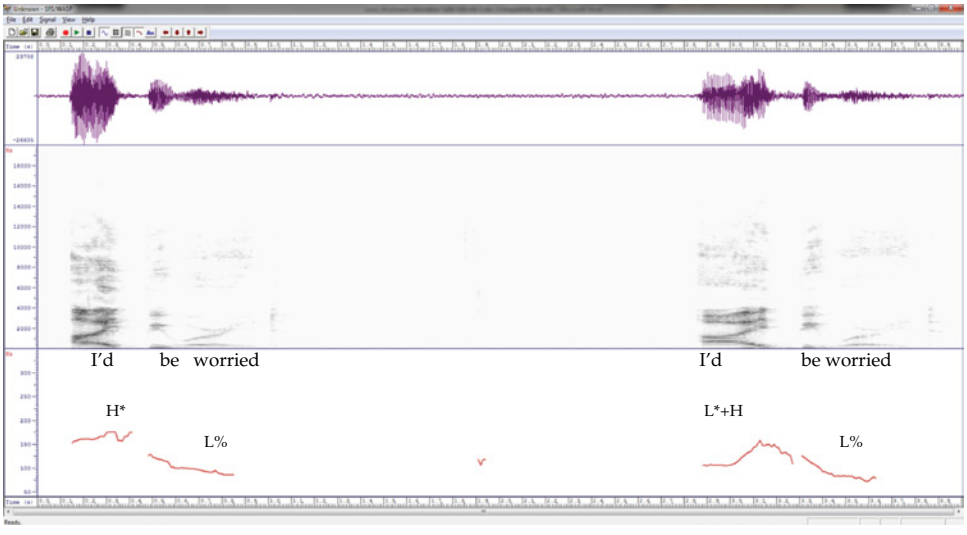


Figure 8.2 H* and L*+H pitch accents on *I'd*.

declines across an utterance unless it is fully reset (Pierrehumbert 1980). An extra-high reset may also be connected to topic shifts (Levis and Pickering 2004; Wichmann 2000). In this case the expanded pitch range is most likely to be the result of contrastive stress – emphasizing *I'd*, presumably in contrast to someone else who is worried. The default position for a nucleus is the last lexical item of a tone group, and this means that the neutral pronunciation of *I'd be worried* would be to have the accent on *worried*. Here, however, the accent has been shifted back to *I'd* and *worried* has been de-accented. This has the effect of signaling that "being worried" is already given in that context and that the focus is on *I'd*. This is a contrastive use of accent and a de-accenting of given or shared information. The contrastive stress may also contain an affective element – emphasis is hard to separate from additional emotional engagement. However, we know that perceptions of emotion are not a function of intonation alone (Levis 1999).

In summary, linguistic uses of intonation in English include:

1. The use of pitch helping to mark juncture between phrases.
2. Pitch accents marking syllables as informationally important.
3. De-accenting syllables following the final pitch accent. This marks information as informationally unimportant.
4. Final pitch movement at the ends of phrases providing general meanings of openness or closedness of content of speech. These include pitch movement at the ends of intermediate and final phrases in an utterance.
5. Extremes of pitch range marking topic shifts or parenthetical information.

Applications in applied linguistics

Acoustic analysis has been used to examine in fine phonetic detail some of the prosodic differences between languages, with important implications for both clinical studies and also the study of second language acquisition (e.g., Mennen 2007). An example of cross-linguistic comparison is the work on pitch range. It is commonly claimed that languages differ in the degree to which their speakers exploit pitch range and that these differences are cultural (e.g., Van Bezooijen 1995), in some cases giving rise to national stereotypes. However, such observations are often drawn from studies whose ways of measuring are not necessarily comparable. Mennen, Schaeffler, and Docherty (2012) examined the pitch range differences between German and English using a variety of measures and found that global measures of the F0 range, as used in many other studies, were less significant than measures based on linguistically motivated points in the contour. The claim, therefore, is not just a reflection of cultural differences but that "f0 range is influenced by the phonological and/or phonetic conventions of the language being spoken" (2012: 2258). Such findings have important implications for L2 acquisition: while some learners may indeed resist an F0 range if it does not accord with their cultural identity, the results of this study show that such cross-language differences in F0 range may also arise from difficulty in acquiring the intonational structure of the other language (2012). This may also be the case for disordered speech, where pitch range has been thought to be symptomatic of certain conditions. Here, too, it may be a case of inadequate mastery of the intonation system rather than its phonetic implementation.

Another area of study made possible by instrumental techniques is the close analysis of the timing of pitch contours in relation to segmental material. Subtle differences in F0 alignment have been found to characterize cross-linguistic prosodic differences. Modern Greek and Dutch, for example, have similar rising (prenuclear) contours but they are timed differently in relation to the segmental material (Mennen 2004). These differences are not always acquired by non-native speakers and contribute, along with segmental differences, to the perception of a foreign accent. Variation in alignment can also be discourse-related. Topic-initial high pitch peaks often occur later in the accented syllable – even to the extent of occurring beyond the vowel segment (Wichmann, House and Rietveld 2000), and our perception of topic shift in a spoken narrative may therefore be influenced not only by pitch height but also by fine differences in peak timing.

Experimental methods are now widely used to examine the interface between phonology and phonetic realization. This includes studies of timing and alignment, as described above, studies to establish the discreteness of prosodic categories underlying the natural variation in production, and also investigations into the phonetic correlates of perceived prominence in various languages. Such experiments sometimes use synthesized stimuli, whose variation is experimentally controlled, and sometimes they rely on specially chosen sentences being read aloud. Laboratory phonology, as it is called, has been seen as an attempt at rapprochement between phonologists who deal only with symbolic representation with no

reference to the physics of speech, and the engineers who use signal- processing techniques with no reference to linguistic categories or functions (Kohler 2006). There are many, however, who criticize the frequent use of isolated, noncontextualized sentences, with disregard for "functions, contextualisation and semantic as well as pragmatic plausibility" (Kohler 2006: 124). Those who study prosody from a conversation analysis perspective have been particularly critical of these methods, rejecting both the carefully designed but unnatural stimuli and also the post hoc perceptions of listeners. The only valid evidence for prosodic function in their view is the behavior of the participants themselves. The experimental approach therefore, however carefully designed and however expertly the results are analysed, has from their perspective little or nothing to say about human interaction.

Intonation and meaning

By exploiting the prosodic resources of English it is possible to convey a wide range of meanings. Some meanings are conveyed by gradient, paralinguistic effects, such as changes in loudness, tempo, and pitch register. Others exploit categorical phenomena including phrasing (i.e., the placement of prosodic boundaries), the choice of tonal contour such as a rise or a fall, and the location of pitch accents or nuclear tones. It is uncontroversial that intonation in English and other intonation languages does not convey propositional meaning: a word has the same "dictionary" meaning regardless of the way it is said, but there is less general agreement on how many other kinds of meaning can be conveyed. Most lists include attitudinal, pragmatic, grammatical, and discoursal meanings, and it is these that will be examined here.

Attitudinal meaning

We know intuitively that intonation can convey emotions and attitudes, but what is more difficult is to ascertain how this is achieved. We should first distinguish between attitude and emotion: there have been many studies of the effect of emotion on the voice, but the accurate recognition of discrete emotions on the basis of voice alone is unreliable. According to Pittham and Scherer (1993), anger and sadness have the highest rate of identification, while others, e.g., fear, happiness, and disgust, are far less easily recognized. If individual emotions are hard to identify, it suggests that they do not have consistent effects on the voice. However, it does seem to be possible to identify certain dimensions of emotion in the voice, namely whether it is active or passive, or whether it is positive or negative (Cowie et al. 2000). As with emotions, there are similar difficulties with identifying attitudes. There is a plethora of attitudinal labels – all familiar from dialogue in fiction (*Yes, he said grumpily; No, she said rather condescendingly*…, etc.). Yet experiments (e.g. Crystal 1969), show that listeners fail to ascribe labels to speech samples with more than a minimum of agreement. O'Connor and Arnold (1973) and Pike (1945)

attempted for pedagogical reasons to identify the attitudinal meanings of specific melodic patterns, but one only has to change the words of the sample utterance to evoke an entirely different meaning. This suggests that the meaning, however intuitively plausible, does not lie in the melodic pattern itself. According to Ladd (1996), the elements of intonation have very general, fairly abstract meaning, but these meanings are "part of a system with a rich interpretative pragmatics, which gives rise to very specific and often quite vivid nuances in specific contexts" (1996: 39–40). In other words, we need to look to pragmatics, and the inferential process, to explain many of the attitudes that listeners perceive in someone's "tone of voice".

Pragmatic meaning

In order to explain some of these pragmatic effects we first need an idea of general, abstract meanings, which, in certain contexts, are capable of generating prosodic implicatures. The most pervasive is the meaning ascribed to final pitch contours: it has been suggested, for example, that a rising tone (L*H) indicates openness (e.g. Cruttenden 1986) or nonfinality (e.g. Wichmann 2000), while falling contours (H*L) indicate closure, or finality. This accounts for the fact that statements generally end low, questions often end high, and also that nonfinal tone groups in a longer utterance also frequently end high, signaling that there is more to come. We see the contribution of the final contour in English question tags; they either assume confirmation, with a falling tone as in: *You've eaten, \haven't you*, or seek information, with a rising tone, as in: *You've eaten, /haven't you?* The "open"–"closed" distinction also operates in the context of speech acts such as requests: in a corpus-based study of *please*-requests, of the type '*Can/could you … please*' (Wichmann 2004), some requests were found to end in a rise, i.e., with a high terminal, and some in a fall, i.e., with a low terminal. The low terminal occurred in contexts where the addressee has no option but to comply and was closer to a polite command, while the rising version, ending high, sounded more tentative or "open" and was closer to a question (consistent with the interrogative form). The appropriateness of each type depends, of course, on the power relationship between the speaker and hearer. If, for example, the speaker's assumptions were not shared by the addressee, a "command", however polite, would not be well received and would lead the hearer to infer a negative "attitude".

The low/high distinction has been said to have an ethological basis – derived from animal signaling where a high pitch is "small" and a low pitch is "big", and, by extension, powerless and vulnerable or powerful and assertive respectively. This is the basis of the Frequency Code proposed by Ohala (1994) and extended by Gussenhoven (2004), who suggests that these associations have become grammaticalized into the rising contours of questions and the falling contours statements, but also underlie the more general association of low with "authoritative" and high with "unassertive".

Another contour that has pragmatic potential is the fall-rise, common in British English. It is frequently exploited for pragmatic purposes to imply some kind of

reservation, and is referred to by Wells (2006: 27–32) as the "implicational fall-rise". In the following exchanges there is an unspoken "but" in each reply, which leaves the hearer to infer an unspoken reservation:

Did you like the film?
The \/acting was good.

Are you using the car?
\/No

How's he getting on at school?
He en\/joys it.

Information structure

It is not only the choice of tonal contour – i.e., whether rising, falling, or falling-rising – but also its location that conveys important information. The placement of prominence in English conveys the information structure of an utterance, in other words how the information it contains is packaged by the speaker in relation to what the hearer already knows. Nuclear prominence is used to focus on what is new in the utterance and the default position is on the last lexical word of a phrase, or more strictly on the stressed syllable of that word. This default placement is the background against which speakers can use prominence strategically to shift focus from one part of an utterance to another. In most varieties of English, the degree of prominence relates to the degree of salience to be given to the word in which it occurs. If the final lexical item is not given prominence it is being treated as given information or common ground that is already accessible to the hearer, and the new information is signaled by prominence elsewhere in the phrase or utterance. In this way, the hearer can be pointed to different foci, often implying some kind of contrast. In the following exchange, the item "money" is being treated as given, but the word "lend" (probably with an implicational fall-rise) sets up an implied contrast with "give":

Can you give me some money?
Well, I can lend you some money.

This technique of indicating what is assumed to be given information or common ground is, like other aspects of intonation, a rich source of pragmatic inference.

Grammatical meaning

A further source of intonational meaning is the phrasing or grouping of speech units through the placement of intonation boundaries (IPs, tone-group boundaries). Phrasing indicates a degree of relatedness between the component parts, whether in terms of grammar, e.g., phrase structure, or mental representations (Chafe 1994). The syntax-intonation mapping is less transparent in spontaneous

speech, but when written text is read aloud, phrase boundaries tend to coincide with grammatical boundaries, and the way in which young children read aloud gives us some insight into their processing of grammatical structures: when word by word reading (*she – was – sitting – in – the – garden*) changes to phrase by phrase (*she was sitting – in the garden*) we know that the reader has understood how words group to become phrases. Phrasing can in some cases have a disambiguating function, as in the difference between ||*He washed and fed the dog*|| and ||*He washed | and fed the dog*|| (Gut 2009). However, such ambiguities arise rarely – they are often cited as examples of the grammatical function of intonation, but in practice it is usually context that disambiguates and the role of intonation is minimal. One important point to note is that pauses and phrase boundaries do not necessarily co-occur. In scripted speech, there is a high probability that any pause is likely to co-occur with a boundary, but not that each boundary will be marked by a pause. In spontaneous speech, pauses are an unreliable indicator of phrasing, since they are performance-related.[4]

Discourse meaning

Texts do not consist of a series of unrelated utterances: they are linked in a variety of ways to create a larger, coherent whole. Macrostructures are signaled *inter alia* by the use of conjunctions, sentence adverbials, and discourse markers. There are also many typographical features of written texts that guide the reader, including paragraphs, headings, punctuation, capitalization, and font changes, all of which provide visual information that is absent when listening to a text read aloud. In the absence of visual information, readers have to signal text structures prosodically, and to do this they often exploit gradient phenomena including pitch range, tempo, and loudness. Pitch range, for example, is exploited to indicate the rhetorical relationships between successive utterances. "Beginnings", i.e., new topics or major shifts in a narrative, often coinciding with printed paragraphs, tend to be indicated by an extra-high pitch on the first accented syllable of the new topic (see Wichmann 2000). In scripted speech this is likely to be preceded by a pause, but in spontaneous monologue there may be no intervening pause but a sudden acceleration of speech into the new topic, the so-called "rush-through" (Couper-Kuhlen and Ford 2004: 9; Local and Walker 2004). In conversation this allows a speaker to change to a new topic without losing the floor to another speaker.

If, in contrast, speakers wish to indicate a strong cohesive relationship between two successive utterances, the pitch range at the start of the second utterance is compressed so that the first accented syllable is markedly lower than expected. Expansion and compression of pitch range also play a part in signaling parenthetical sequences. Typically these are lower in pitch and slightly faster than the surrounding speech, but sometimes there is a marked expansion instead; in each case the parenthetical utterance is marked out as "different" from the main text (Dehé and Kavalova 2007). These prosodic strategies for marking macrostructures are also observable in conversational interaction, where they are combined with many more subtle phonetic signals that, in particular, enable the

management of interaction in real time. This aspect of discourse prosody is the focus of much work in the CA framework (see Chapter 11 by Szczepek Reed in this volume).

The strategies described above are all related to the structure of the text itself, but there is a recent strand of research into intonational meaning that investigates how the pitch relationships across speaker turns can signal interpersonal meaning, such as degrees of rapport between speakers. This is seen as an example of a widely observed mirroring, or accommodation, between conversational participants. It occurs in many ways – in posture, gesture, accent, and in prosody. Meaning is made not by any inherent characteristics of individual utterances or turns at speaking but by the sequential patterning, in other words, how an utterance relates prosodically to that of another speaker. We know, for example, that the timing of response particles such as *mhm*, *right*, *ok*, etc., is important: if they are rhythmically integrated with the rhythm of the other speaker they are perceived to be supportive, while a disruption of the rhythm is a sign of disaffiliation (Müller 1996). Sequential pitch matching has similarly been found to be a sign of cooperativeness or affiliation: speakers tend to accommodate their pitch register to that of their interlocutor over the course of a conversation (Kousidis et al. 2009) and this engenders, or reflects, rapport between the speakers.[5] This can have an important effect on interpersonal relations, as has been observed in a classroom setting (Roth and Tobin 2009): "in classes … where we observe alignment in prosody, participants report feeling a sense of solidarity …"(2009: 808).

In summary, at a very abstract level, the placement and choice of pitch accents and the way in which speech is phrased can convey grammatical and pragmatic meaning, such as speech acts, and also the information structure of an utterance. The phonetic realization of these choices, i.e., exploiting the range of highs and lows that the voice can produce, can convey discourse discontinuities, such as paragraphs or topic shifts, and, conversely, continuities in the cohesive relations between successive utterances. All these choices can also be exploited by speakers to generate pragmatic implicatures, which are often interpreted as speaker "attitudes". Finally, the whole range of prosodic resources, including pitch, loudness, and timing are drawn on to manage conversational interaction, including ceding and taking turns, competing for turns, and holding the floor, and in the creation or expression of interpersonal rapport.

NOTES

1 Anyone wanting to carry out a laboratory study of, say, final pitch contours, might thus be unwise to devise sentences to be read aloud ending in words such as *hush, sack, stretch*. Easier to study would be sentences ending in *roam, lane, ring*, etc.

2 For example, in order to indicate contrast or information structure as given information tends to be less prominent than new information.

3 The sentence was recorded by one of the authors using WASP (University College London's Speech, Hearing and Phonetic Sciences Division's computer program, http://www.phon.ucl.ac.uk/resource/sfs/wasp.htm).
4 This is possibly too categorical – there is much more to be said about pauses, including their orientation to the hearer (e.g., see Clark 1996).
5 It is not always clear whether the rapport between speakers leads to accommodation or whether the accommodation leads to rapport. This is a big topic, i.e., whether discourse reflects social relations or is constitutive of them. The author is not sure whether to go into this.

REFERENCES

Beckman, M.E. and Pierrehumbert, J.B. 1986. Intonational structure in Japanese and English. *Phonology* 3(1): 255–309.

Bing, J.M. 1980. *Aspects of English Prosody*, Bloomington, IN: Indiana University Linguistics Club.

Boersma, P. 2001. Praat: a system for doing phonetics by computer. *Glot International* 5(9/10): 341–345.

Bradford, B. 1988. *Intonation in Context*, Cambridge: Cambridge University Press.

Brazil, D. 1994. *I*, Cambridge: Cambridge University Press.

Celce-Murcia, M., Brinton, D., and Goodwin, J. 2010. *Teaching Pronunciation: A Course Book and Reference Guide*, New York: Cambridge University Press.

Chafe, W. 1994. *Discourse, Consciousness, and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing*, Chicago: The University of Chicago Press.

Clark, H.H. 1996. *Using language*, Cambridge: Cambridge University Press.

Couper-Kuhlen, E. and Ford, C.E. 2004. *Sound Patterns in Interaction*, Amsterdam: John Benjamins.

Cowie, R., Douglas-Cowie, E., Savvidou, S., McMahon, E., Sawey, M., and Schröder, M. 2000. Feeltrace: an instrument for recording perceived emotion in real time. In: *Speech and Emotion: Proceedings of the ISCA Workshop*, R. Cowie, E. Douglas-Cowie, and M. Schröder (eds.), 19–24, Belfast NI: Textflow.

Cruttenden, A. 1986. *Intonation*, Cambridge: Cambridge University Press.

Cruttenden, A. 1997. *Intonation*, 2nd edition, Cambridge: Cambridge University Press.

Crystal, D. 1969. *Prosodic Systems and Intonation in English*, Cambridge: Cambridge University Press.

Dehé, N. and Kavalova, Y. (eds.). 2007. *Parentheticals*, Amsterdam: John Benjamins Publishing.

Grice, M.(2006. *Intonation*. In: *Encyclopedia of Language and Linguistics*, K. Brown (ed.), 2nd edition, 778–788, Oxford: Elsevier.

Gussenhoven, C. 2004. *The Phonology of Tone and Intonation*, Cambridge: Cambridge University Press.

Gut, U. 2009. *Non-native Speech: A Corpus Based Analysis of Phonological and Phonetic Properties of L2 English and German*, Frankfurt am Main: Peter Lang.

Hahn, L.D. 2004. Primary stress and intelligibility: research to motivate the teaching of suprasegmentals. *TESOL Quarterly* 38(2): 201–223.

Halliday, M.A.K. 1967. *Intonation and Grammar in British English*, The Hague: Mouton.

Kenworthy, J. 1987. *Teaching English Pronunciation*, New York: Longman.

Kohler, K. 2006. Paradigms in experimental prosodic analysis: from measurement to function. In: *Language, Context and Cognition: Methods in Empirical Prosody Research*, S. Sudhoff (ed.), Berlin: De Gruyter.

Kousidis, S., Dorran, D., McDonnell, C., and Coyle, E. 2009. Time series analysis of acoustic feature convergence in human dialogues. In: *Digital Media Centre Conference Papers*, Dublin Institute of Technology.

Ladd, D.R. 1996. *Intonational Phonology*, Cambridge: Cambridge University Press.

Ladd, D.R. 2008. *Intonational Phonology*, 2nd edition, Cambridge: Cambridge University Press.

Levis, J.M. 1999. Intonation in theory and practice, revisited. *TESOL Quarterly* 33(1): 37–63.

Levis, J.M. and Grant, L. 2003. Integrating pronunciation into ESL/EFL classrooms. *TESOL Journal* 12(2), 13–19.

Levis, J. and Pickering, L. 2004. Teaching intonation in discourse using speech visualization technology. *System* 32(4): 505–524.

Local, J. and Walker, G. 2004. Abrupt joins as a resource for the production of multi-unit, multi-action turns. *Journal of Pragmatics* 36(8): 1375–1403.

Mennen, I. 2004. Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of Phonetics* 32(4): 543–563.

Mennen, I. 2007. Phonological and phonetic influences in non-native intonation. In: *Non-native Prosody: Phonetic Descriptions and Teaching Practice*, J. Trouvain and U. Gut (eds.), 53–76, The Hague, Mouton De Gruyter

Mennen, I., Schaeffler, F., and Docherty, G. 2012. Cross-language difference in F0 range: a comparative study of English and German. *Journal of the Acoustical Society of America* 131(3): 2249–2260.

Müller, F. 1996. Affiliating and disaffiliating with continuers: prosodic aspects of recipiency. In: *Prosody in Conversation*, E. Couper-Kuhlen and M. Selting (eds.), 131–176, New York: Cambridge University Press.

O'Connor, J.D. and Arnold, G.F. 1973. *Intonation of Colloquial English*,2nd edition, London: Longman.

Ohala, J.J. 1994. The frequency code underlies the sound-symbolic use of voice pitch. In: *Sound Symbolism*, L. Hinton, J. Nichols, and J.J. Ohala (eds.), 325–347, Cambridge: Cambridge University Press.

Pickering, L. 2001. The role of tone choice in improving ITA communication in the classroom. *TESOL Quarterly* 35(2), 233–255.

Pierrehumbert, J.B. 1980. The phonology and phonetics of English intonation. Doctoral dissertation, Massachusetts Institute of Technology.

Pierrehumbert, J. and Hirschberg, J. 1990. The meaning of intonational contours in the interpretation of discourse. In: *Intentions in Communication*, P. Cohen, J. Morgan, and M. Pollack (eds.), 271–310, Cambridge, MA: MIT Press.

Pike, K. 1945. The intonation of American English. Ann Arbor, MI: University of Michigan Press.

Pittham, J. and Scherer, K.R. 1993. Vocal expression and communication of emotion. *Handbook of Emotions*, New York: Guilford Press.

Roth, W.-M. Tobin, K. 2010. Solidarity and conflict: prosody as a transactional resource in intra- and intercultural communication involving power differences. *Cultural Studies of Science Education* 5: 805–847.

Schmerling, S. 1976. *Aspects of English Sentence* Stress. Austin, TX: University of Texas Press.

Steele, J. 1775. *An Essay Towards Establishing the Melody and Measure of Speech*, London: Bowyer & Nichols [Reprinted, Menston, The Scholar Press, 1969].

Van Bezooijen, R. 1995. Sociocultural aspects of pitch differences between Japanese and Dutch women. *Language and Speech* 38: 253–265.

Wells, J. 2006. *English Intonation*, Cambridge: Cambridge University Press.

Wennerstrom, A. 1994. Intonational meaning in English discourse: a study of non-native speakers. *Applied Linguistics*,15(4): 399–420.

Wennerstrom, A. 1998. Intonation as cohesion in academic discourse. *Studies in Second Language Acquisition* 20(1): 1–25.

Wennerstrom, A. 2001. *The Music of Everyday Speech: Prosody and Discourse Analysis*, Oxford: Oxford University Press.
Wichmann, A. 2000. *Intonation in Text and Discourse*, London: Longman.
Wichmann, A. 2004. The intonation of *please*-requests. *Journal of Pragmatics* 36: 1521–1549.
Wichmann, A., House, J., and Rietveld, T. 2000. Discourse effects on f0 peak timing in English. In: *Intonation: Analysis, Modelling and Technology*, A. Botinis (ed.), 163–182, Dordrecht: Kluwer Academic Publishers.
Williams, B. 1996. The formulation of an intonation transcription system for British English. In: *Working with Speech*, G. Knowles, A. Wichmann, and P. Alderson (eds.), 38–57, London: Longman.

Part III Pronunciation and Discourse

9 Connected Speech

GHINWA ALAMEEN AND JOHN M. LEVIS

Introduction

Words spoken in context (in connected speech) often sound quite different from those same words when they are spoken in isolation (in their citation forms or dictionary pronunciations). The pronunciation of words in connected speech may leave vowel and consonant sounds relatively intact, as in some types of linking, or connected speech may result in modifications to pronunciation that are quite dramatic, including deletions, additions, or changes of sounds into other sounds, or combinations of all three in a given word in context. These kinds of connected speech processes (CSPs) are important in a number of areas, including speech recognition software, text-to-speech systems, and in teaching English to second language learners. Nonetheless, connected speech, in which segmental and suprasegmental features interact strongly, lags far behind work in other areas of segmentals and suprasegmentals in second language research and teaching. Some researchers have argued that understanding CSPs may be particularly important for the development of listening skills (Field 2008; Jenkins 2000; Walker 2010), while others see CSPs' production as being particularly important for more intelligible pronunciation (Celce-Murcia et al. 2010; Reed and Michaud 2005).

Once a word is spoken next to other words, the way it is pronounced is subject to a wide variety of processes. The changes may derive from linguistic context (e.g., *can be* said as *cam be*), from speech rate (e.g., tomorrow's *temperature* runs from 40 in the morning to 90 at midday, in which temperature may be said as tɛmpɹətʃɚ, tɛmpətʃɚ, or tɛmtʃɚ, depending on speed of speech), or from register (e.g., *I don't know* spoken with almost indistinct vowels and consonants but a distinctive intonation in very casual speech). When these conditioning factors occur together in normal spoken discourse, the changes to citation forms can become cumulative and dramatic.

Connected speech processes based on register may lead to what Cauldwell (2013) calls jungle listening. Just as plants may grow in isolation (in individual

The Handbook of English Pronunciation, First Edition. Edited by Marnie Reed and John M. Levis.

pots in a greenhouse), they may also grow in the company of many other plants in the wild. The same is true of words. Typically, the more casual and informal the speech register is, the more the citation forms of words may change. As a result, the pronunciation of connected speech may become a significant challenge to intelligibility, both the intelligibility of native speech for non-native listeners and the intelligibility of non-native speech for native listeners. Connected speech, perhaps more than other features of English pronunciation, demonstrates the importance of intelligibility in listening comprehension. In many elements of English pronunciation, non-native speakers need to speak in a way that is intelligible to their listeners, but connected speech processes make clear that non-native listeners must also learn to understand the speech of native words that may sound quite different from what they have come to expect, and their listening ability must be flexible enough to adjust to a range of variation based not only on their interlocutors but also on the formality of the speech.

Definitions of connected speech

Hieke (1987) defined connected speech processes as "the changes which conventional word forms undergo due to the temporal and articulatory constraints upon spontaneous, casual speech" (1987: 41). That is, they are the processes that words undergo when their border sounds are blended with neighboring sounds (Lass 1984). Citation form pronunciations occur in isolated words under heavy stress or in sentences delivered in a slow, careful style. By contrast, connected speech forms often undergo a variety of modifications that cannot always be predicted by applying phonological rules (Anderson-Hsieh, Riney, and Koehler 1994; Lass 1984; Temperley 1987). It may be that all languages have some form of connected speech processes, as Pinker (1995: 159–160) claims:

> In speech sound waves, one word runs into the next seamlessly; there are no little silences between spoken words the way there are white spaces between written words. We simply hallucinate word boundaries when we reach the edge of a stretch of sound that matches some entry in our mental dictionary. This becomes apparent when we listen to speech in a foreign language: it is impossible to tell where one word ends and the next begins.

Although CSPs are sometimes thought to be a result of sloppy speech, they are completely normal (Celce-Murcia et al. 2010; Henrichsen 1984). Highly literate speakers tend to make less use of some CSPs (Prator and Robinett 1985); however, even in formal situations, such processes are completely acceptable, natural, and essential part of speech.

Similar modifications to pronunciation also occur within words (e.g., *input* pronounced as *imput*), but word-based modifications are not connected speech since they are characteristic pronunciations of words based on linguistic context alone (the [n] moves toward [m] in anticipation of the bilabial stop [p]). In this chapter, we will not address changes within words but only those between words.

Function of CSPs in English

The primary function of CSPs in English is to promote the regularity of English rhythm by compressing syllables between stressed elements and facilitating their articulation so that regular running speech timing can be maintained (Clark and Yallop 1995). For example, certain closed class words such as prepositions, pronouns, and conjunctions are rarely stressed, and thus appear in a weak form in unstressed contexts. Consequently, they are "reduced" in a variety of processes to preserve the rhythm of the language. Reducing speech can also be attributed to the law of economy where speakers economize on effort, avoiding, for example, difficult consonant sequences by eliding sounds (Field 2003). The organs of speech, instead of taking a new position for every sound, tend to connect sounds together using the same or intermediate articulatory gestures to save time and energy (Clarey and Dixson 1963).

One problem that is noticeable in work on connected speech is the types of features that are included in the overall term. Both the names given to the connected speech processes and the phenomena included in connected speech vary widely in research and in ESL/EFL textbooks. Not only are the types and frequency of processes dependent on rhythmic constraints, speech register, and linguistic environment, the types of connected speech processes may vary among different varieties of English.

A classification for connected speech processes

In discussing connected speech, two issues cannot be overlooked: differences in terminology and the infrequency of relevant research. Not only do different researchers and material designers use different terms for CSPs (e.g., sandhi variations, reduced forms, absorption), they also do not always agree on how to classify them. In addition, conducting experimental studies of connected speech can be intimidating to researchers because "variables are normally not controllable and one can never predict the number of tokens of a particular process one is going to elicit, which in turn makes the application of statistical measures difficult or impossible" (Shockey 2003: 109). As a result, only a few people have researched CSPs in relation to English language teaching and have done so only sporadically (Brown and Kondo-Brown 2006).

Connected speech terminology varies widely, as does the classification of the CSPs. This is especially true in language teaching materials, with features such as contractions, blends (coalescent assimilation or palatalization), reductions (unstressed words or syllables), linking, assimilation (progressive and regressive), dissimilation, deletion (syncope, apocope, aphesis), epenthesis flapping, disappearing /t/, *gonna/wanna* type changes, –s and –ed allomorphs, and linking. This small selection of terms suggests that there is a need for clarity in terminology and in classification.

We propose that connected speech processes be classified into six main categories: linking, deletion, insertion, modification, reduction, and multiple processes.

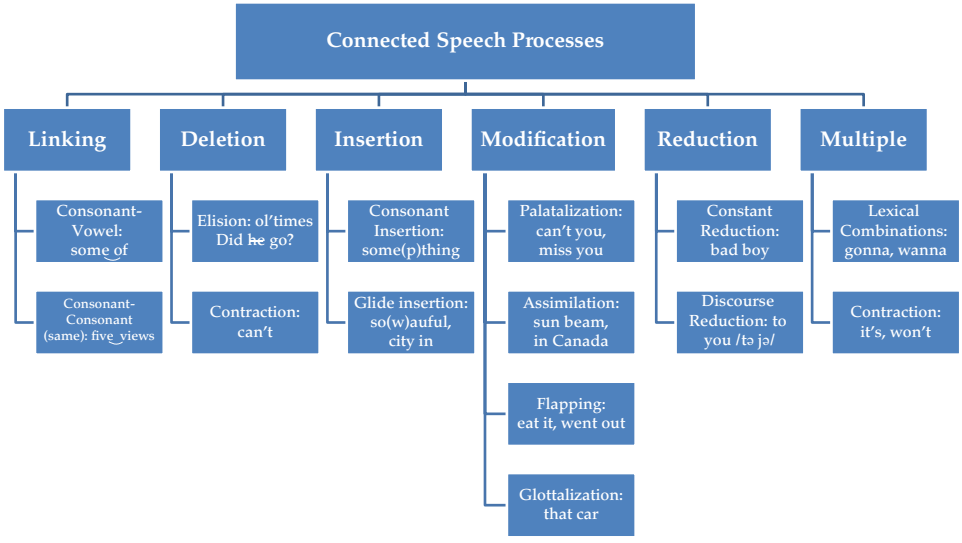


Figure 9.1 Our categorization of Connected Speech Processes.

Our proposed chart is in Figure 9.1. Linking, the first category, is the only one that does not involve changes to the segments of the words. Its function in connected speech is to make two words sound like one without changes in segmental identity, as in the phrases *some_of* [sʌm əv] and *miss_Sarah* [mɪs sɛɹə]. Linking can result in resyllabification of the segments without changing them [sʌ.məv] or in lengthening of the linked segments in cases where both segments are identical, e.g., [mɪsːɛɹə]. Our description of linking is narrower than that used by many writers. We restrict linking to situations in which the ending sound of one word joins the initial sound of the next (a common enough occurrence), but only when there is no change in the character of the segments. Other types of links include changes, and we include them in different categories. For example, the /t/ in the phrase *hat band* would be realized as a glottal stop and lose its identity as a [t], i.e., [hæʔbæ̃nd]. We classify this under our category of modifications. In addition, in the phrase *so awful*, the linking [w] glide noticeably adds a segment to the pronunciation, i.e., [soʷɔfəɬ]. We classify this under additions.

The second category, deletion, involves changes in which sounds are lost. Deletions are common in connected speech, such as potential loss of the second vowel in a phrase like *see it* [siːt] in some types of casual speech, the loss of [h] in pronouns, determiners, and auxiliaries (e.g., *Did ~~he~~ do ~~his~~ ~~h~~omework?, Their friends ~~h~~ave already left*) or deletions of medial consonant sounds in complex consonant groupings (e.g., *the bes~~t~~ gift, old~~d~~ times*). Some types of contractions are included in the category, mainly where one or more sounds are deleted in a contraction (e.g., *cannot* becomes *can't*).

The third category, insertion, involves changes that add sounds. An example would be the use of glides to combine two vowels across words (e.g., Popeye's

statement of *I am what I am* → *I yam what I yam*). Consonant additions also occur, as in the intrusive /r/ that is characteristic of some types of British or British-influenced English (*The idea of* → *The idea(r) of*). There are few insertions of vowels across word boundaries, although vowel insertion occurs at the lexical level, as in *athlete* → *athelete* as spoken by some NAmE speakers.

The fourth category is modification. Changes involve modifications to pronunciation that substitute one phoneme for others (e.g., *did you* pronounced as [dɪdʒu] rather than [dɪdju], or less commonly, modifications that are phonetically (allophonically) but not phonemically distinct (e.g., *can you* pronounced as [kɛŋju] rather than [kɛnju]). The palatalization examples are more salient than changes that reflect allophonic variation. Other examples of modifications include assimilation of place, manner, or voicing (e.g., *on point*, where the /n/ becomes [m] before the bilabial stop); flapping (*sit̲ around* or *wen̲t outside*, in which the alveolar stops or nasal-stop clusters are frequently pronounced as alveolar oral or nasal flaps in NAmE); and glottalization, in which /t/ before nasals or stops are pronounced with a distinct glottal articulation (*can't make it, that car* as [kænʔmekɪt] and [ðæʔkɑɹ]).

The fifth category is reduction. Reductions primarily involve vowels in English. Just as reduced vowels are lexically associated with unstressed syllables, so words may have reduced vowels when spoken in discourse, especially word classes such as one-syllable determiners, pronouns, prepositions, and auxiliaries. Reductions may also involve consonants, such as the lack of release on stop consonants as with the /d/ in a phrase like *bad boy*, for some speakers.

The final category, multiple CSPs, involves instances of lexical combination. These are highly salient lexical chunks that are known for exhibiting multiple CSPs in each lexical combination. These include chunks like *gonna* (*going to* in full form), with its changes of [ŋ] to [n], vowel reduction in *to*, modifications of the [o] to [ʌ] in *going*, and the deletion of the [t]. Other examples of lexical combinations are *What do you/What are you* (both potentially realized as *whatcha/whaddya*) and *wanna* (for *want to*). In addition, we also include some types of contractions in this category, such as *they're*, *you're, it's*, and *won't*. All three of these involve not only deletions but modifications such as vowel changes and voicing assimilation.

The final category points out a common feature of CSPs. The extent to which phonetic form of authentic utterances differs from what might be expected is illustrated by Shockey (2003). That is, the various types of CSPs occur together, not only in idiomatic lexical combinations but also in all kinds of language. This potentially makes connected speech sound very different from citation forms of the same lexical items. For example, the phrase *part of* is subject to both flapping and linking, so that its phonetic quality will be [pʰɑɹ.ɾəv].

Connected speech features

It appears that certain social and linguistic factors affect the frequency, quality, and contexts of CSPs. Lass (1984) attributes CSPs to the immediate phonemic environment, speech rate, the formality of the speech situation, and other social factors, such as

social distance. Most researchers distinguish two styles of speech: casual everyday style and careful speech used for certain formal occasions, such as presentations. According to Hieke (1984), in casual spontaneous speech, speakers pay less attention to fully articulating their words, hence reducing the distinctive features of sounds while connecting them. Similarly, when examining linking for a native speaker (NS) and a non-native speaker (NNS) of English, Anderson-Hsieh, Riney, and Koehler (1994) found that style shifting influenced the manner in which speakers link their words. In their study, NSs and NNSs performed more linking in spontaneous speech tasks than those involving more formal sentence reading.

However, other studies have found that while there was some evidence that read speech was less reduced, unscripted and scripted speech shows great phonological similarity (Alameen 2007; Shockey 1974). The same processes apply to both styles and nearly to the same degree. Native speakers do not seem to know that they are producing speech that differs from citation form. In Alameen (2007), NNSs as well as NSs of English did not have significant differences between their linking performance in text reading and spontaneous speech tasks, which indicates that a change in speech style may not entail a change in linking frequency. Furthermore, Shockey (2003) noted that many CSPs occur in fast speech as well as in slow speech, so "if you say 'eggs and bacon' slowly, you will probably still pronounce 'and' as [m], because it is conventional – that is, your output is being determined by habit rather than by speed or inertia" (2003: 13).

Other factors, such as social distance, play a role in determining the frequency with which such processes happen (Anderson-Hsieh, Riney and Koehler 1994). When the speaker and the listener both belong to the same social group and share similar speech conventions, the comprehension load on the listeners will be reduced, allowing them to pay less attention to distinctive articulation.

Variation in degree is another feature that characterizes CSPs. Many researchers tend to think of connected speech processes in clear-cut definitions; however, speakers do not always produce a specific CSP in the same way. A large study of CSPs was done at the University of Cambridge, results of which appeared in a series of articles (e.g., Barry 1984; Wright 1986). The results showed that most CSPs produce a continuum rather than a binary output. For instance, if the process of contraction suggests that *do not* should be reduced to *don't*; we often find, phonetically, cases of both expected variations and a rainbow of intermediate stages, some of which cannot be easily detected by ear. Such findings are insightful for CSP instruction since they help researchers and teachers decide on what CSP to give priority to depending on the purpose and speech style. They also provide a better understanding of CSPs that may facilitate the development of CSP instructional materials.

Research into CSPs

Various studies have investigated an array of connected speech processes in native speaker production and attempted to quantify their characteristics. These studies examined processes such as assimilation and palatalization (Barry 1991; Shi et al.

2005), deletion (R.W. Norris 1994), contraction (Scheibman 2000), British English liaison (Allerton 2000), linking (Alameen 2007; Hieke 1987; Temperley 1987), and nasalization (Cohn 1993). Such studies provide indispensable background for any research in L2 perception and pronunciation. The next sections will look at studies that investigated the perception and production of NNSs connected speech in more detail.

Perception

The perception of connected speech is closely connected to research on listening comprehension. In spoken language, frustrating misunderstandings in communication may arise because NSs do not pronounce English the way L2 learners are taught in the classroom. L2 learners' inability to decipher foreign speech comes from the fact that they develop their listening skills based on the adapted English speaking styles they experience in an EFL class. In addition, they are often unaware of the differences between citation forms and modifications in connected speech (Shockey 2003). When listening to authentic L2 materials, Brown (1990: 4) claims an L2 learner:

> Will hear an overall sound envelope with moments of greater and lesser prominence and will have to learn to make intelligent guesses, from all the clues available to him, about what the probable content of the message was and to revise this interpretation if necessary as one sentence follows another – in short, he has to learn to listen like a native speaker.

A part of the L2 listener's problem can be attributed to the fact that listening instruction has tended to emphasize the development of top-down listening processes over bottom-up processes (Field 2003; Vandergrift 2004). However, in the past decade, researchers have increasingly recognized the importance of bottom-up skills, including CSPs, for successful listening (Rost 2006). In the first and only book dedicated to researching CSPs in language teaching, Brown and Kondo-Brown (2006) note that, despite the importance of CSPs for learners, little research on their instruction has been done, and state that the goal of their book is to "kick-start interest in systematically teaching and researching connected speech" (2006: 6). There also seems to be a recent parallel interest in CSPs studies in EFL contexts, especially in Taiwan (e.g., Kuo 2009; Lee 2012; Wang 2005) and Japan (e.g., Crawford 2006; Matsuzawa 2006). The next section will discuss strategies NSs and NNSs use to understand connected speech, highlight the effect of CSPs on L2 listening and review the literature on the effectiveness of CSPs perceptual training on listening perception and comprehension.

Speech segmentation

A good place to start addressing L2 learners' CSPs problems is by asking how native listeners manage to allocate word boundaries and successfully segment speech. Some models of speech perception propose that specific acoustic markers

are used to segment the stream of speech (e.g., Nakatani and Dukes 1977). In other models, listeners are able to segment connected speech through the identification of lexical items (McClelland and Elman 1986; D. Norris 1994). Other cues to segmentation can also be triggered by knowledge of the statistical structure of lexical items in the language in the domains of phonology (Brent and Cartwright 1996) and metrical stress (Cutler and Norris 1988; Grosjean and Gee 1987). In connected speech, the listener compares a representation of the actual speech stream to stored representations of words. Here, the presence of CSPs may create lexical ambiguity due to the mismatch between the lexical segments and their modified phonetic properties. For experienced listeners, however, predictable variation does not cause a breakdown in perception (Gaskell, Hare, and Marslen-Wilson 1995).

On the other hand, several speech perception models have been postulated to account for how L2 listeners segment speech. Most focus on the influence of the L1 phonological system on L2 perception, for example, the Speech Learning Model (Flege 1995), the Perceptual Assimilation Model (Best 1995), and the Native Language Magnet Model (Kuhl 2000). In order to decipher connected speech, NNSs depend heavily on syntactic-semantic information, taking in a relatively large amount of spoken language to process. This method introduces a processing lag instead of processing language as it comes in (Shockey 2003). L2 learners' speech segmentation is primarily led by lexical cues pertaining to the relative usage frequency of the target words, and secondarily from phonotactic cues pertaining to the alignment of syllable and word boundaries inside the carrier strings (Sinor 2006). This difference in strategy leads to greater difficulty in processing connected speech because of the relatively less efficient use of lexical cues.

CSPs in perception and comprehension

The influence of connected speech on listening perception (i.e., listening for accuracy) and comprehension (i.e., listening for content) has been investigated in several studies (Brown and Hilferty 1986; Henrichsen 1984; Ito 2006). These studies also show how reduced forms in connected speech can interfere with listening comprehension. Evidence that phoneme and word recognition are indeed a major source of difficulty for low-level L2 listeners comes from a study by Goh (2000). Out of ten problems reported by second language listeners in interviews, five were concerned with perceptual processing. Low-level learners were found to have markedly more difficulties of this kind than more advanced ones.

In a pioneer study in CSP research, Henrichsen (1984) examined the effect of the presence and absence of CSPs on ESL learners' listening comprehension skills. He administered two dictation tests to NNS of low and high proficiency levels and NSs. The results confirmed his hypothesis that reduced forms in listening input would decrease the saliency of the words and therefore make comprehension more difficult for ESL learners. Comprehending the input with reduced forms, compared to when the sentences were fully enunciated, was more difficult for both levels of students, indicating that connected speech was not easy to understand regardless of the level of the students.

Ito (2006) further explored the issue by adding two more variables to Henrichsen's design: modification of sentence complexity in the dictation test and different types of CSPs. She distinguished between two types of reduced forms, lexical and phonological forms. Her assumption was that *lexical reduced forms* (e.g., *won't*) exhibit more saliency and thus would be more comprehensible compared to *phonological forms* (e.g., *she's*). As in Henrichsen's study, the non-native participants scored statistically significantly higher on the dictation test when reduced forms were absent than when they were present. Furthermore, NNSs scored significantly lower on the dictation test of phonological forms than that of lexical forms, which indicated that different types of reduced forms did distinctively affect comprehension. Considering the effects of CSPs on listening perception and comprehension and the fact that approximately 35% of all words can be reduced in normal speech (Bowen 1975), perceptual training should not be considered a luxury in the language classroom.

Effectiveness of CSP training on perception and comprehension

Since reduced forms in connected speech cause difficulties in listening perception and comprehension, several research studies have attempted to investigate the effectiveness of explicit instruction of connected speech on listening. After Henrichesen's findings that features of CS reduced perceptual saliency and affected ESL listeners' perception, other researchers have explored the effectiveness of teaching CS to a variety of participants. In addition to investigating whether L2 perceptual training can improve learners' perceptual accuracy of CSPs, some of the researchers examined the extent to which such training can result in improved overall listening comprehension (Brown and Hilferty 1986; Carreira 2008; Lee and Kuo 2010; Wang 2005). The types of CSPs that could be taught effectively with perceptual training or which are more difficult for students were also considered in some studies (Crawford 2006; Kuo 2009; Ting and Kuo 2012). Furthermore, students' attitudes toward listening difficulties, types of reduced forms, and reduced forms instruction were surveyed (Carreira 2008; Kuo 2009; Matsuzawa 2006).

The range of connected speech processes explored in those studies was not comprehensive. Some focused on teaching specific high-frequency modifications, i.e., word combinations undergoing various CSPs and appearing more often in casual speech than others; for instance *gonna* for *going to*, palatalization in *couldja* instead of *could you* (Brown and Hilferty 1986; Carreira 2008; Crawford 2006; Matsuzawa 2006). Others researched certain processes, such as C-V linking, palatalization, and assimilation (Kuo 2009; Ting and Kuo 2012). These studies trained participants to recognize CSP general rules using a great number of reduction examples, instead of focusing on a limited number of examples and teaching them repeatedly.

Results of the previous studies generally indicate that CSP instruction facilitated learners' perception of connected speech. However, most studies failed to address

the long-term effects of such training on learners' perceptual accuracy. Moreover, no study has investigated generalization and transfer of improvement to novel contexts, which indicates that improved abilities could extend beyond the training to natural language usage.

Production

Connected speech is undeniably important for perception, but it is also important for production. Most language teaching materials emphasize exercises meant to teach L2 learners how to pronounce connected speech features more successfully, based on the assertion that "these guidelines will help your comprehension as well as your pronunciation of English" (Grant 1993: 157). Temperley (1987) suggests that "closer examination of linking shows its more profound effect on English pronunciation than is usually recognized, and that its neglect leads to misrepresentation and unnatural expectations" (1987: 65). However, the study of connected speech phenomena has been marginalized within the field of speech production. This section discusses connected speech production in NS and NNS speech, highlighting its significance and prevalence, and demonstrating the effectiveness of training in teaching CS production.

CSPs in production

Hieke (1984, 1987), Anderson-Hsieh, Riney, and Kochear (1994), and Alameen (2007) investigated aspects of connected speech production of American English, including linking, and compared them to those of non-native speakers of English. In a series of studies, Hieke (1984, 1987) investigated the prevalence and distribution of selected CSPs in native and non-native speech. Samples of spontaneous, casual speech were collected from NS (n = 12) and NNS (n = 29) participants according to the paraphrase mode, that is, they retold a story heard just once. C-V linking, alveolar flapping, and consonant cluster reduction were considered representative of major connected speech categories in these studies. Hieke (1987) concluded that these phenomena could be considered "prominent markers of running speech" since they "occur in native speech with sufficient consistency to be considered regular features of fluency" (1987: 54).

Building on Hieke's research, Anderson-Hsieh, Riney, and Kochler (1994) examined linking, flapping, vowel reduction, and deletion, in the English of Japanese ESL learners, comparing them to NSs of American English. The authors examined the production of intermediate-proficiency (IP) and high-proficiency (HP) NNSs by exploring the extent to which style-shifting affected the CSPs of ESL learners. Results showed that while the HP group approximated the performance of the native speaker group, the IP group often lagged far behind. An analysis of the reduced forms used revealed that the IP group showed a strong tendency to keep word boundaries intact by inserting a glottal stop before the word-initial vowel in the second word. The HP group showed the same tendency but less frequently.

Alameen (2007) replicated Anderson-Hsieh et al.'s (1994) macroanalytical study while focusing on only C-V and V-V linking. Results indicated that beginning-proficiency and intermediate-proficiency participants linked their words significantly less often than NS participants. However, the linking rates of the two NNS groups were similar despite the difference in proficiency level. While supporting past research findings on linking frequency, results of the study contradicted Anderson-Hsieh et al.'s (1994) results in terms of finding no significant difference between spontaneous and reading speech styles. In addition, the study showed that native speakers linked more frequently towards function words than to content words.

Effectiveness of CSP training on production

Although there have been numerous studies on the effectiveness of teaching CSP on listening perception and comprehension, very little research has been conducted on CSP production. This can be largely attributed to the pedagogical priorities of teaching listening to ESL learners since they are more likely to listen than to speak in ESL contexts and partly to a general belief that CSPs are not a central topic in pronunciation teaching and sometimes markers of "sloppy speech". Three research studies (Kuo 2009; Melenca 2001; Sardegna 2011) have investigated the effectiveness of CSP instruction on L2 learners. Interestingly, all studies were primarily interested in linking, and all were masters or PhD theses. This can probably be accounted for by the facts that (a) linking, especially C-V linking, is the simplest and "mildest" CSP (Hieke 1987) since word boundaries are left almost intact, (b) linking as a phenomenon is prevalent in all speech styles, while other CSPs are more frequent in more informal styles, e.g., palatalization, and (c) L2 problems in linking production can render production disconnected and choppy and, hence, difficult for NS to understand (Dauer 1992) and unlinked speech can sometimes be viewed as aggressive and abrupt (Anderson-Hsieh, Riney, and Kochler 1994; Hatch 1992).

Melenca (2001) explored the influence of explicitly teaching Japanese speakers of English how to connect speech so as to avoid a robotic speech rhythm. A control ($N = 4$) and an experimental group ($N = 5$) were each given three one-hour sessions in English. Their ability to link word pairs was rated using reading aloud and elicited free-speech monologues that were compared to an NS baseline. Descriptive statistics showed that individual performances in pre- and post-test varied considerably. Yet they also demonstrated that the performance of experimental group participants either improved or remained relatively stable in linking ability while the CG performance stayed the same. Noteworthy are the findings that the average percentages of linking while reading a text was at 67% and while speaking freely at 73%. This suggests that linking occurs with approximately equal frequency under both conditions. Melenca, furthermore, recommended that C-V and V-V linking be taught in one type of experiment, while C-C linking should be investigated in a separate study, due to the variety and complexity of C-C linking contexts.

By training EFL elementary school students in Taiwan on features of linking for 14 weeks, Kuo (2009) examined whether such training positively affected students' speech production. After receiving instruction, the experimental group significantly improved their speech production and developed phonological awareness. Among the taught categories, V-V linking posed more problems for the experimental group due to its high degree of variance.

In spite of the positive influence of training measured immediately after the treatment, effectiveness of the training cannot be fully evaluated without examining the long-term effects of such training. Sardegna (2011) attempted to fill this gap. Using the Covert Rehearsal Model (Dickerson 1994), she trained 38 international graduate students on how to improve their ability to link sounds within and across words. A read-aloud test was administered and recorded twice during the course, and again five months to two years after the course ended. The results suggested that students maintained a significant improvement over time regardless of their native language, gender, and length of stay in the United States prior to instruction. However, other learner characteristics and factors seemed to contribute to greater or lesser improvement over time, namely (a) entering proficiency level with linking, (b) degree of improvement with linking during the course, (c) quantity, quality, and frequency of practice with linking when using the covert rehearsal model, (d) strong motivations to improve, and (e) prioritization of linking over other targets for focused practice.

The studies show that CSP training can help NNSs improve their speech production both immediately after the treatment and in delayed post-tests. More importantly, the previous studies reveal several problem areas on which researchers need to focus in order to optimize time spent in researching CSP production training. A longer period of instruction may facilitate more successful output. Practising several types of CSPs can be time-consuming and confusing to students (Melenca 2001). Finally, there is a need for exploring newer approaches to teaching CSPs that could prove to be beneficial to L2 learners.

Future research into connected speech

A more complete understanding of connected speech processes is essential for a wide variety of applications, from speech recognition to text-to-speech applications to language teaching. In English language teaching, which we have focused on in this chapter, CSPs have already been the focus of heavy attention in textbooks, much of which is only weakly grounded in research. There is a great need to connect the teaching of CSPs with research. Although we have focused on research that is connected to applied linguistics and language teaching, this is not the only place that research is being done. Speech recognition research, in particular, could be important for pedagogy in the need to provide automated feedback on production.

Previous studies suggest several promising paths for research into CSPs. The first involves the effects of training and questions about classroom priorities. It is generally agreed that intelligibility is a more realistic goal for language learners

than is native-like acquisition (Munro and Derwing 1995). In addition, intelligibility is important both for acquisition of perception and for acquisition of production (Levis 2005). Most language teaching materials today include exercises on CSPs without clear priorities about which CSPs are most important. Is linking more important for spoken intelligibility than addressing insertion or deletion? We also know that CSPs can improve with training, but we do not know whether improvement increases intelligibility. Since practising many types of CSPs during the same training period can be confusing to students, CSPs that are likely to make the greatest difference should be emphasized in instruction.

Next, it is not clear if there is an optimal period of training for improvement. A longer period of instruction may facilitate more successful learning. In addition, we do not know which type of input is optimal. CSPs occur in both read and spontaneous speech, formal and informal, and for some types of CSPs there is very little difference in frequency of occurrence for both ways of speaking (Alameen 2007; Melenca 2001). The reading task approximates the spontaneous speech task in actual linking levels. It remains to be seen as to whether using read speech is best for all CSPs, or whether different types of input may serve different purposes, including raising awareness, improving perception, or improving production.

Thirdly, there is a need for exploring newer approaches to teaching CSPs that could prove beneficial to L2 learners, especially the use of electronic visual feedback (EVF). Coniam (2002) demonstrated that EVF can be valuable in raising awareness of stress-timed rhythm. Alameen (2014) demonstrated that the same kind of awareness can be developed for linking. Since pronunciation time is limited in any classroom, EVF is a promising way to promote autonomous learning of CSPs outside the classroom.

CSPs are among the most diverse, complex, and fascinating phonological phenomena, and despite inconsistent research on them, are deserving of greater attention. While these features of speech are likely to be universal, they are also language specific in how they are realized. While research into CSPs is not abundant in English, it is far less abundant for other languages. French is an exception to this rule, with research into liaison. Spanish synalepha is another documented type of CSP, but other languages have no body of research to speak of. This means that there is also a great need for research into CSPs in other languages.

REFERENCES

Alameen, G. 2007. The use of linking by native and non-native speakers of American English. Unpublished MA thesis, Iowa State University, Ames, IA.

Alameen, G. 2014. The effectiveness of linking instruction on NNSs speech perception and production. Unpublished doctoral dissertation, Iowa State University, Ames, IA.

Allerton, D. 2000. Articulatory inertia vs "systemzwang": changes in liaison phenomena in recent British English. *English Studies* 6: 574–581.

Anderson-Hsieh, J., Riney, T., and Koehler, K. 1994. Connected speech modifications in the English of Japanese ESL learners. *IDEAL* 7: 31–52.

Barry, M. 1984. Connected speech: processes, motivation, models. *Cambridge Papers in Phonetics and Experimental Linguistics* 3.

Barry, M. 1991. Assimilation and palatalisation in connected speech. Presented at *The ESCA Workshop on Phonetics and Phonology of Speaking Styles*, Barcelona, Spain, 1–9.

Best, C. 1995. A direct realist view of cross-language speech perception. In: *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, W. Strange (ed.), 171–204, Timonium, MA: York Press.

Bowen, J.D. 1975. *Patterns of English Pronunciation*, Rowley, MA: Newbury House.

Brent, M. R. and Cartwright, T.A. 1996. Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition* 61: 93–125.

Brown, G. 1990. *Listening to Spoken English*, 2nd edition, London and New York: Longman.

Brown, J.D. and Hilferty, A. 1986. The effectiveness of teaching reduced forms for listening comprehension. *RELC Journal* 17: 59–70.

Brown, J.D. and Kondo-Brown, K. 2006. Introducing connected speech. In: *Perspectives on Teaching Connected Speech to Second Language Speakers*, J.D. Brown and K. Kondo-Brown (eds.), 1–15), Manoa and Honolulu, HI: National Foreign Language Resource Center, University of Hawaii at Manoa.

Carreira, J.M. 2008. Effect of teaching reduced forms in a university preparatory course. In: *JALT2007 Conference Proceedings*, K. Bradford-Watts, T. Muller, and M.S. Swanson (eds.), 200–207, Tokyo: JALT.

Cauldwell, R. 2013. *Phonology for Listening: Teaching the stream of speech*. Birmingham: speechinaction.

Celce-Murcia, M., Brinton, D.M., Goodwin, J.M., and Griner, B. 2010. *Teaching Pronunciation Paperback with Audio CDs (2): A Course Book and Reference Guide*, 2nd edition, Cambridge University Press.

Clarey, M.E. and Dixson, R.J. 1963. *Pronunciation Exercises in English*, New York: Regents.

Clark, J. and Yallop, C. 1995. *An Introduction to Phonetics and Phonology*, Oxford: Blackwell.

Cohn, A.C. 1993. Nasalisation in English: phonology or phonetics. *Phonology* 10(1): 43–81.

Coniam, D. 2002. Technology as an awareness-raising tool for sensitising teachers to features of stress and rhythm in English. *Language Awareness* 11(1): 30–42.

Crawford, M.J. 2006. A study on teaching reductions perceptually. In: *JALT 2005 Conference Proceedings*, K. Bradford-Watts, C. Ikeguchi, and M. Swanson (eds.), Tokyo: JALT.

Cutler, A. and Norris, D. 1988. The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance* 14: 113–121.

Dauer, R.M. 1992. *Accurate English: A Complete Course in Pronunciation*, Prentice-Hall.

Dickerson, W.B. 1994. Empowering students with predictive skills. In: *Pronunciation Pedagogy and Theory: New Views, New Directions*, J. Morley (ed.), 17–33, Alexandria, VA: TESOL Publications.

Field, J. 2003. Promoting perception: lexical segmentation in L2 listening. *ELT Journal* 57(4): 325–334.

Field, J. 2008. Bricks or mortar: Which parts of the input does a second language listener rely on? *TESOL Quarterly* 42(3): 411–432.

Flege, J.E. 1995. Second language speech learning: theory, findings, and problems. In: *Speech Perception and Linguistic*

Experience: Issues in Cross-language Research, W. Strange (ed.), 233–276, Timonium, MA: York Press.
Gaskell, M.G., Hare, M., and Marslen-Wilson, W.D. 1995. A connectionist model of phonological representation in speech perception. *Cognitive Science* 19: 407–439.
Goh, C.C.M. 2000. A cognitive perspective on language learners' listening comprehension problems. *System* 28: 55–75.
Grosjean, F. and Gee, J.P. (1987). Prosodic structure and spoken word recognition. *Cognition* 25: 135–156.
Hatch, E.M. 1992. *Discourse and Language Education*, Cambridge: Cambridge University Press.
Henrichsen, L.E. 1984. Sandhi-variation: a filter of input for learners of ESL. *Language Learning* 34(3): 103–123.
Hieke, A.E. 1984. Linking as a marker of fluent speech. *Language and Speech* 27: 343–354.
Hieke, A.E. 1987. Absorption and fluency in native and non-native casual speech in English. In: *Sound Patterns in Second Language Acquisition*, A. James and J. Leather (eds.), Dordrecht, The Netherlands and Providence, RI: Foris.
Ito, Y. 2006. The significance of reduced forms in L2 pedagogy. In *Perspectives on Teaching Connected Speech to Second Language Speakers*, J.D. Brown and K. Kondo-Brown (eds.), 17–26, Manoa, Honolulu, HI: National Foreign Language Resource Center, University of Hawaii at Manoa.
Jenkins, J. 2000. *The Phonology of English as an International Language*, 1st edition, Oxford and New York: Oxford University Press.
Kuhl, P. 2000. A new view of language acquisition. *Proceedings of the National Academy of Science* 97(22): 11850–11857.
Kuo, H.C. 2009. The effect of English linking instruction on EFL elementary school students' speech production and phonological awareness. Unpublished MA thesis, National Chung Cheng University, Chiayi, Taiwan.
Lass, R. 1984. *Phonology*, Cambridge: Cambridge University Press.
Lee, J.-T. 2012. A comparative study on the effectiveness of communicative and explicit connected speech instruction on Taiwanese EFL junior high school students' listening comprehension. Unpublished MA thesis, National Chunghua University of Education, Taiwan.
Lee, J.-T. and Kuo, F.-L. 2010. Effects of teaching connected speech on listening comprehension. In: *Selected Papers from the Nineteenth Symposium on Englsih Teaching*, 153–162.
Levis, J.M. 2005. Changing contexts and shifting paradigms in pronunciation teaching. *Tesol Quarterly* 39(3): 369–377.
Matsuzawa, T. 2006. Comprehension of English reduced forms by Japanese business people and the effectiveness of instruction. In: *Perspectives on Teaching Connected Speech to Second Language Speakers*, J.D. Brown and K. Kondo-Brown (eds.), 59–66, Manoa, Honolulu, HI: National Foreign Language Resource Center, University of Hawaii at Manoa.
McClelland, J.L. and Elman, J.L. 1986. The TRACE model of speech perception. *Cognitive Psychology* 18: 1–86.
Melenca, M.A. 2001. Teaching connected speech rules to Japanese speakers of English so as to avoid a staccato speech rhythm. Unpublished thesis, Concordia University.
Munro, M.J. and Derwing, T.M. 1995. Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning* 45(1): 73–97.
Nakatani, L. H. and Dukes, K.D. 1977. Locus of segmental cues for word juncture. *Journal of the Acoustical Society of America* 62: 715–719.
Norris, D. 1994. Shortlist: a connectionist model of continuous speech recognition. *Cognition* 52: 189–234.

Norris, R.W. 1994. Keeping up with native speaker speed: an investigation of reduced forms and deletions in informal spoken English. *Studies in Comparative Culture* 25: 72–79.

Pinker, S. 1995. *The Language Instinct: How the Mind Creates Language*, New York: HarperPerennial.

Prator, C. H. and Robinett, B.W. 1985. *Manual of American English Pronunciation*, New York: Holt, Rinehart, and Winston.

Reed, M. and Michaud, C. 2005. *Sound Concepts: An Integrated Pronunciation Course*, New York: McGraw-Hill.

Rost, M. 2006. Areas of research that influence L2 listening instruction. In: *Current Trends in the Development and Teaching of the Four Language Skills*, E. Usó Juan and A. Martínez Flor (eds.), 47–74, Berlin and New York: M. de Gruyter.

Sardegna, V.G. 2011. Pronunciation learning strategies that improve ESL learners' linking. In: *Proceedings of the 2nd Pronunciation in Second Language Learning and Teaching Conference*, J. Levis and K. LeVelle (eds.), 105–121, Ames, IA: Iowa State University.

Scheibman, J. 2000. I dunno: a usage-based account of the phonological reduction of don't in American English conversation. *Journal of Pragmatics* 32: 105–124.

Shi, R., Gick, B., Kanwischer, D., and Wilson, I. 2005. Frequency and category factors in the reduction and assimilation of function words: EPG and acoustic measures. *Journal of Psycholinguistic Research* 34(4): 341–364.

Shockey, L. 1974. Phonetic and phonological properties of connected speech. *Ohio State Working Papers in Linguistics* 17: 1–143.

Shockey, L. 2003. *Sound Patterns of Spoken English*, Malden, MA: Blackwell Publishing.

Sinor, M. 2006. Lexical and phonotactic cues to speech segmentation in a second language. Unpublished Doctoral dissertation, University of Alberta.

Temperley, M.S. 1987. Linking and deletion in final consonant clusters. In: *Current Perspectives on Pronunciation: Practices Anchored in Theory*, J. Morley (ed.), Teachers of English to Speakers of Other Languages.

Ting, W.-Y. and Kuo, F.-L. 2012. Messages behind the unheard sounds: crossing the word boundaries through songs. *NCUE Journal of Humanities* 5: 75–92.

Vandergrift, L. 2004. Listening to learn or learning to listen? *Annual Review of Applied Linguistics* 24: 3–25.

Walker, R. 2010. *Teaching the Pronunciation of English as a Lingua Franca*, Pap/Com edition, Oxford and New York: Oxford University Press.

Wang, Y.T. 2005. An exploration of the effects of reduced forms instruction on EFL college students' listening comprehension. Unpublished MA thesis, National Tsing Hua University, Hsinchu, Taiwan.

Wright, S. 1986. The interaction of sociolinguistic and phonetically-conditioned CPSs in Cambridge English: auditory and electropalatographic evidence. *Cambridge Papers in Phonetics and Experimental Linguistics* 5.

10 Functions of Intonation in Discourse

ANNE WICHMANN

Introduction

Prosody is an integral part of the spoken language. It conveys structure and meaning in an individual utterance, and it also contributes to the structuring and meaning of discourse. It is this latter aspect that is increasingly being seen as an important dimension of language learning. According to Levis and Pickering (2004: 506), there is "growing recognition that traditional sentence-level approaches may not be able to meet the needs of language teachers and learners". Indeed, there are several studies, as reported in Piske (2012), which suggest that "learners profit to a larger extent from instruction that focuses on suprasegmental aspects of pronunciation" (2012: 54). The purpose of this chapter is therefore to outline some of the ways in which prosody, and intonation in particular, serves to structure spoken texts, manage interaction, and convey pragmatic meaning.

Theoretical and methodological frameworks

There are different approaches to the study of prosody and the results are often contradictory. Prosody research is driven not only by different theories of language and human interaction but also by different goals. Early studies, especially in the nineteenth century and before, focused on speech as performance. Speaking was thought of as an art, a rhetorical skill that was crucial for success in politics, in the Church, and in the theatre. A crucial part of the art was known as modulation – described in impressionistic terms, with little clear indication of what the speaker should actually do, other than to "establish a sympathy" with the audience (Brewer 1912: 83). More recent twentieth century analyses of English intonation were pedagogical in focus, driven by the needs of non-native rather than native speakers; this pedagogical tradition persists, for example, in the work of John Wells (2006),

The Handbook of English Pronunciation, First Edition. Edited by Marnie Reed and John M. Levis.

and is clearly of continued importance wherever English is being learnt as a second or foreign language.

In recent decades, with advances in technology, a new motivation for speech research has emerged. This is the desire to design computers that can synthesize human-sounding speech and also understand human speech. Applications of such work are, of course, limited to certain styles of speech: spoken monologue ("reading aloud" continuous text) and goal-oriented dialogue, such as service encounters. Casual conversation, on the other hand, is the main focus of work in interactional linguistics (derived from conversation analysis), with especial interest in how conversation is managed by the participants, reflecting the fundamentally cooperative nature of human communication.

For each of these approaches to discourse prosody there is a range of phonetic features that are thought to be important. Early voice-training manuals refer impressionistically to pace, pitch, and loudness in rather global terms, as the properties of stretches of discourse. The British pedagogical literature, on the other hand, and the British system of intonation in general, describes intonation (and it is usually only intonation and not the other prosodic components) in terms of localized contours – holistic movements such as fall, rise, and fall-rise. These pitch movements are the property of accented syllables and associated unstressed syllables, and it is the choice of contour, its placement, and its phonetic realization that makes an important contribution to discoursal and pragmatic meaning. The American autosegmental system describes the same local pitch movements, not in terms of holistic contours but in terms of their component pitch targets. Thus a rising contour is decomposed into a low target point followed by a high target point, and what is perceived holistically as a rising contour is the interpolation of pitch between those two points. The autosegmental theory of intonation (Pierrehumbert 1987; Pierrehumbert and Hirschberg 1990) has become the standard in most areas of prosody research. In addition, however, the advances in signal processing and the automatic analysis of the speech signal mean that there is a renewed interest in more "global" features, i.e., phonetic features that are the property of longer stretches of speech. These include the average pitch of an utterance or sequence and also long-term variation in tempo and amplitude.

Speakers clearly have a wide range of prosodic resources at their disposal: pitch, loudness, tempo, and voice quality, and can exploit them in various ways. Misunderstandings or loss of intelligibility can arise from errors related to both the phonological inventory and its phonetic implementation, and from choices at both local and global levels. Research in all of these many areas and in a variety of theoretical frameworks therefore has the potential to reveal how we use prosody, and thus raise awareness of its importance among teachers and learners.

Sentence types and speech acts

Although native speakers are rarely conscious of the intonational choices they make, they can certainly tell if something is unusual and does not correspond to what they perceive to be the norm. This can be illustrated by a high-profile pattern

in current use. Over the last 30 years a pattern of intonation has been spreading in English that is a source of great annoyance to older speakers (always a good indication of language change!). This is often called "uptalk" and refers to the use of a rising contour at the end of a statement, instead of the expected falling contour.

The fact that this innovation is so controversial tells us something about the default intonation contours relating to different kinds of sentence types (specifically statements and questions). The traditional pedagogical literature on English intonation makes simple claims about canonical forms: statements and Wh-questions terminate in a falling contour while a yes-no question terminates in a rise. Wh-questions can be used with a rise, but then have a softening, sometimes patronizing, effect. The validity of these claims is sharply contested by those who study conversation from an interactional (conversation-analytic) perspective, but they form useful bases, not only in teaching and in clinical contexts (see, for example, Peppé and McCann 2003) but especially in experimental and large-scale corpus studies geared towards improving speech technology.

While human beings generally have no great difficulty in assessing what a speaker intends with a given utterance – statement, request, greeting, etc. – machines are less adept at doing this. Much research effort has been, and continues to be, invested into modeling human speech (production and recognition) in order to develop speech technology. This includes speech synthesis, automatic speech recognition, and human–machine interaction systems. Any utterances that can only be understood in context pose a challenge to automatic analysis. Shriberg et al. (1998) found it particularly difficult to distinguish automatically between backchannels, e.g., *uhuh*, and agreements, e.g., *yeah*, particularly since some of these are lexically ambiguous (see the section below on backchannel). It was found that agreements had a higher energy level than backchannels, which is assumed to be because emotional involvement is greater in an agreement than a simple "continuer", and greater emotional involvement involves greater energy and often higher pitch. However, the attempt to disambiguate presupposes a single function for each utterance, although linguists have shown that speech acts can be multifunctional: an "agreement" might well also function as a "backchannel".

A good example of multifunctionality is the act of thanking. Aijmer (1996) shows that thanking goes beyond the expression of gratitude. It can be dismissive (e.g., *I can do it myself thank you*), ironic (*thank you, that's all I needed*), and can also initiate a closing sequence, acting simultaneously as an expression of gratitude and a discourse organizer. According to Aijmer, gratitude is expressed differently and has a different intonational realization, depending on the size of the favor: */thank you* with a rising tone sounds casual – it is used in situations where the "favor" is minimal, as in buying a train ticket (e.g., A: *Here you are*. B: */Thank you*). Where more gratitude is being expressed, a falling tone is used (e.g., A: *I'll look after the children for the day if you lik*e. B: *Oh that's so kind of you – \thank you)* (see also Archer et al. 2012: 263 for further examples). This is consistent with the view of Wells (2006: 66), who suggests that the difference between using a rising and falling tone is the difference between "routine acknowledgment" (/thank you) and "genuine gratitude" (\thank you).

The pragmatic consequences of different intonational realizations of the same utterance are to be seen in Wichmann (2004) in a corpus-based study of *please*-requests. Such requests occurred with either a falling or a (falling-)rising final contour. For example: *Can you open the \door please* versus *Can you open the \door / please* (i.e., with final rise on *please*). It was found that those requests with a falling contour were generally used in asymmetrical situations, such as service encounters, where the imposition was socially licensed, while the rising contour was used where compliance could not be taken for granted. Thus, a request at a ticket office for "*a ticket to \Lancaster, please*" assumes that the hearer's role is to comply. On the other hand, a request to borrow something from a friend generally does not make that assumption, and "*could I borrow your |pen /please*" would be more likely. These "default" realizations can, of course, be used strategically regardless of the context: a falling tone might be used to sound "assertive", while the more tentative rising tone might be used to express politeness by suggesting that the hearer has an option to refuse even if it is not actually the case. In other words, such patterns can be used to create the symmetry or asymmetry desired by the speaker, and not just as a reflection of existing relationships. However, if used unwittingly, these choices can also be the source of misunderstandings, particularly in conversation with a native speaker. If the "assertive" version is used innocently in a situation where the speaker does not have the right to demand compliance, it can cause offence. Similarly, a casual-sounding *thank you* (with a rising tone) might offend a hearer who believes that greater gratitude should be expressed. Whether these pragmatic inferences are likely to be drawn in conversation between NNS (i.e., in English as a lingua franca situation) is a matter for future research.

Information structure

A feature of some varieties of English and other Germanic languages is that they use patterns of weak and strong (stressed) syllables to structure the speech in a rhythmic way, both at word level and at utterance level (known as stress-timed rhythm). Vowel quality depends on stress patterns: unstressed syllables tend to be realized with a schwa, or reduced even further, while an accented syllable will contain a full vowel. Deterding (2012: 21) claims that a syllable-timed rhythm (with consequent absence of reduced syllables) may actually enhance intelligibility, and an insistence that learners acquire a stress-based rhythm may be inappropriate.

This may be true in relation to word stress, which is not part of prosody but part of the lexicon – information that is to be found in a dictionary. Sentence stress, on the other hand, is manipulated by the speaker, and is strongly related to the structuring of information in discourse. Processing is no longer a matter of word recognition but of understanding "the flow of thought and the flow of language" (Chafe 1979). The placement of sentence stress reflects what a speaker assumes is in the consciousness of the hearer at the time, and thus is an example of how discourse is co-constructed.

The default position for "sentence"-stress in English is the last potentially stressed syllable in a prosodic group, but this "norm" can be exploited strategically to indicate that an item is already "given" (or accessible in the mind of the hearer). "Givenness" can relate to a single lexical item that has already been referred to: the plain statement, *She's got a KITten*, will have the sentence accent in the default position, namely on the last lexical item. However, in the following exchange, the 'kitten' is given: e.g., *Shall we buy her a KITten? She's already GOT a kitten*. Givenness can also be notional rather than lexical: e.g., *Shall we buy her a KITten? No – she's alLERgic to cats*. Here, the word *cats* subsumes *kitten*: an allergy to adult animals can be taken as including an allergy to kittens.

Research into the brain's response to accentuation patterns has shown that these patterns are important for the hearer in the processing of the ongoing discourse. Baumann and Schumacher (2011) maintain that prosodic prominence (at least in Germanic languages such as English and German) influences the processing of information structure: "information status and prosody have an independent effect on cognitive processing .... More precisely, both newness and deaccentuation require more processing effort (in contrast to givenness and accentuation") (Baumann, personal communication). Similar results have been shown by other researchers. Dahan, Tanenhaus, and Chambers (2002) used eye tracking technology to establish that if an item was accented, the hearer's sight was directed towards non-given items, but towards given items if unaccented. A similar eye-tracking experiment by Chen, Den Os, and De Ruiter (2007) showed that certain pitch contours also biased the listener to given or new entities: a rise-fall strongly biased towards a new entity, while a rise or unaccentedness biased towards givenness.

These experiments might lead one to expect that the deaccentuation of given items is universal. This is not the case – many languages, and some varieties of English (e.g., Indian English, Caribbean English, and some East Asian varieties), do not follow the pattern of Standard British English or General American. It therefore remains to be seen what the processing consequences would be for a speaker of such a language. Taking the production perspective, Levis and Pickering (2004) claim that learners tend to insert too many prominences and that these can obscure the meaning of the discourse. They suggest that practising prominence placement at sentence level, i.e., with no discourse context, might exacerbate this tendency to overaccentuate.

One way of raising awareness of prosodic prominence is to use signal processing software to visualize speech. We know something about the phonetic correlates of perceptual prominence thanks to the seminal work of Fry (1955, 1958). An accented syllable generally displays a marked excursion (upwards or downwards) of pitch, measured as fundamental frequency (F0), together with an increase in duration and amplitude. Cross-linguistic comparisons, such as that carried out by Gordon and Applebaum (2010), provide evidence of the universality of the parameters, even if they are weighted differently in different languages.

Finally, it is important to note that classroom discourse itself may not be the best style of speaking to illustrate the prosody of "given" and "new". In contrast

to most research findings, Riesco Bernier and Romero Trillo (2008) found that in some classroom discourse the distinction between "given" and "new" was not evident in the prosody. However, they chose a very particular kind of discourse: "Let's see/ milk/ does milk come from/ plants/ or animals? Animals/ Animals/ that's right/ from the cow." Although the authors do not say this, it suggests that speaking style in pedagogical situations may in fact be very different from the naturally occurring prosody students are being prepared for.

Text structure

A printed page provides the reader with far more information than the words alone. Typographical conventions, such as punctuation, capitalization, bracketing, and change of font, help the reader to recover the internal structure at the level of the clause and sentence. Paragraph indentation, blank lines, and headings (and subheadings) help the reader to group sequences of sentences into meaningful units. In some kinds of text, bullet points and numbered lists are also an aid to organizing the information on the page. Of course, none of this information is available when a text is read aloud, and the listener is reliant on the reader's voice – pauses and changes in pitch, tempo, and loudness – to indicate the structure of the text.

The idea of "spoken paragraphs" was addressed by Lehiste (1979), who established not only that readers tended to mark these prosodically but also that listeners used the prosodic information to identify the start of a new topic. In read speech, the position of pauses suggests breaks in a narrative, with longer pauses being associated with paragraph breaks. However, the most reliable prosodic correlate of topic shift is a pitch "reset", an increase in pitch range. This observation – that an increase in pitch range accompanies a major shift in the discourse – has been made for both read-aloud speech and spontaneous conversation, and in languages other than English (Brazil, Coulthard, and Johns 1980; Brown, Currie, and Kenworthy 1980; Nakajima and Allen 1993; Yule 1980).

While there is some agreement that the boundaries between units of text are prosodically marked, there is less agreement as to whether there are any internal features that operate across a "paragraph". Sluijter and Terken (1993) claimed that a paragraph was not only marked at its boundaries but that each successive sentence within the paragraph displayed a narrower range. The idea is that there is a kind of "supra-declination" that mirrors the declination (tendency for pitch to gradually fall/the pitch envelope to become narrower) across a single sentence, but at the level of the paragraph. This was certainly true for their experimental data, but is less evident in naturally occurring data, mainly because of many competing discourse effects on pitch range, such as parenthesis, reported speech, and cohesive devices (see Wichmann 2000).

While speakers intuitively use prosodic text-structuring devices in conversation, they do not do so consistently when reading aloud. Their use depends very much of the skill of the reader, and many readers are simply not very skilled. Some

readers such as newsreaders, for example, are highly paid professionals, but experimental studies of read speech sometimes have to rely on readers recruited from the general public or from student groups – whoever is prepared to offer their time. Kong (2004), who looked at topic structuring in Korean, found that her female speakers marked the structure of their spontaneous narratives much more consistently than when they read aloud a subsequent transcription of them.

It is important to remember, however, that paragraph divisions in written texts are typographical conventions, and do not necessarily map on to meaningful text units. Some texts, especially literary texts, have a very fluid topic structure, shifting gradually from one "scene" to the next. Orthographic paragraphs "indicate not the boundary between one clearly definable episode and another, but a point in a text where one or more of the coherent scenes, temporal sequences, character configurations, event sequences and worlds .... change more or less radically (Chafe, 1979: 180) Since much of the research into prosodic text segmentation has been carried out with Automatic Speech Recognition (ASR) in mind, such complex texts are rarely used, and the focus is generally on texts in which orthographic divisions map consistently on to meaningful units.

An awareness of the effective prosodic structuring of spoken discourse, particularly spoken monologue such as lectures, is thought to be important in teaching. Thompson (2003) claims that the awareness of intonational "paragraphs" is as important for understanding lectures as it is for performing them, and the training of lecturers in speaking skills should therefore also include awareness of phonological structuring. She compared five English for Academic Purposes (EAP) training texts (for listening skills) with six authentic undergraduate lectures. In the authentic data she found longer phonological paragraphs but fewer metatextual cues (*first*, *next*, *in conclusion*, etc.). The EAP training texts, on the other hand, appeared to focus on metatextual comment with little reference to phonological structuring. Thompson suggests that students are not well served by these texts and that learning to "hear" the structure of authentic lectures might help them. She concedes that some EAP teachers avoid intonation as "difficult to teach" but suggests that broad topic shifts can be pointed out and consciousness raised without a lot of technical detail about intonation.

Interaction management: turn-taking in conversation

Spontaneous speech displays many of the same structuring devices as prepared speech, including the kind of pitch resets discussed above. If someone is telling a story the shifts in the narrative will be marked prosodically, just as they are in read-aloud speech. There will be some differences, however, depending on whether the speaker is "licensed" to take an extended turn, or whether other speakers are waiting to take a turn at speaking at the first opportunity. A licensed narrative gives the speaker the space to pause and reflect without risking interruption. This is the case in a lecture, for example, or in a media interview. In casual conversation there is an expectation that all participants have equal rights to the floor, and speakers are

especially vulnerable to interruption when they are ending a topic and wanting to start another. Pauses are therefore not reliable topic cues in spontaneous narrative. Speakers frequently launch new topics by omitting a pause and accelerating from the end of one topic into the new. This is known as "rush through" (Couper-Kuhlen and Ford 2004: 9; Local and Walker 2004). It is particularly evident in political interviews, when the interviewee hopes to control the talk and therefore avoid further questions that might raise new and possibly uncomfortable topics.

This is just one of the devices used in the management of turn-taking, which is an important aspect of conversation, and one in which prosody, along with gaze, gesture, and other nonverbal phenomena, plays a part. It is remarkable how smoothly some conversations appear to run, and it has been claimed (Sacks, Schegloff, and Jefferson 1974) that while there is overlap and also silence, there are frequent cases of no-gap-no-overlap, often referred to as "latching". These are of course perceptual terms, and recent acoustic analysis (Heldner 2011) has shown that a gap is not perceived until after a silence of more than 120 ms, a perceived overlap is overlapping speech of more than 120 ms, and no-gap-no-overlap is perceived when the silence or overlap is less than 120 ms. (Wilson and Wilson 2005 had already predicted a less than 200 ms threshold). It seems that smooth turn-taking is less common than has been assumed, applying to fewer than one-fifth of the turns analyzed. However, we cannot assume that any speech overlap at turn exchanges is necessarily an interruption, as Edelsky (1981) showed. Some overlapping speech is intended to support the current speaker, and therefore distinguishes between competitive and collaborative overlap.

The prosodic characteristics of the end of a turn are generally thought to be a lowering of pitch and a slowing down. It is clear, however, that these features alone cannot account for smooth turn-taking nor can they function as reliable cues. Work in the conversation analysis framework (e.g., Szcepek-Reed 2011) finds too little regularity in the shape of turns to justify any generalizations about the prosody of turn-ceding or turn-holding. The smoothness of transition at turn exchanges suggests that participants cannot be waiting for the other speaker to be silent before taking a turn, or even for the final pitch contour, but must have some way of projecting and preparing for an upcoming turn relevant place (TRP) in advance. The cues used in projecting a TRP have been widely discussed (see references in Wilson and Wilson 2005) and include semantic, syntactic, prosodic, body movement-/gaze-related cues. However, as Wilson and Wilson (2005) point out, there may be many cues that indicate an upcoming TRP but which nonetheless do not indicate the exact timing of it. They suggest an alternative, cognitive, account of what appears to be universal behavior, despite some cultural differences. They propose that conversation involves "a fine tuned coordination of timing between speakers" (2005: 958). In other words, the timing of turn-taking is governed by mutual rhythmic entrainment, possibly on the basis of the syllable rate, despite wide variation in syllable length; speakers converge in their speech rate rather like relay runners getting into step before taking the baton. This notion of "entrainment" or "accommodation" as applied to speech will be discussed in more detail in the final section below.

Backchannel

The successful management of conversation depends not only on smooth turn-taking but on the successful elicitation of small responses, sometimes known as "continuers" or "backchannels". A simple test for this is to consciously withhold any verbal or nonverbal response when another person is speaking to you. They will very soon stop and ask what is wrong. Speakers of a second language therefore must not only be intelligible themselves, they must also be able to indicate to an interlocutor the degree to which they are following a conversation.

A very early study (Yngve 1970) referred to short responses as "getting a word in edgewise". The pervasiveness of these responses in conversation is confirmed by Jurafsky et al. (1997) (cited in Ward and Tsukahara 2000) who find that short responses constitute 19% of all utterances in a corpus of American English conversation. Studying short responses, however, is complicated by the number of different words or nonword vocalizations that can be used as a backchannel: Benus, Gravano, and Hirschberg (2007) in their study of American English found in their Games corpus that *mmhm*, *uhhuh*, *okay*, and *yeah* were the most common, followed by *right*, *yes/yep*, and *alright*. While vocalizations such as *mmhm* and *uhhuh* are easily recognizable as backchannels, both *okay* and *yeah* are multifunctional. *Okay*, for example, can be used to signal agreement and to mark a topic shift, in addition to functioning as a backchannel response, although Benus, Gravano, and Hirschberg (2007), in an attempt to disambiguate, found that backchannels have "higher pitch and intensity and greater pitch slope than affirmative words expressing other pragmatic functions" (2007: 1065).

Backchannel responses are not randomly produced, but at points that seem to be cued by the current speaker; in other words, speakers "ask" for backchannel. Ward and Tsukahara (2000) indicate clear evidence that backchannel feedback is cued in most cases by the speaker. A possible cue is a period of low pitch, while Benus, Gravano, and Hirschberg (2007: 1065) identify "phrase-final rising pitch as a salient trigger for backchanneling". This accounts for the interpretation of "uptalk" as a trigger (Hirschberg and Ward 1995). Even these cues, however, do not explain the precision timing of backchannel responses, and Wilson and Wilson's (2005) notion of "entrainment" may offer an explanation here too.

It is important for language learners to know that there are cross-cultural differences in turn-taking behavior, including backchanneling. For example, there are cultural differences in backchannel frequency, and this difference alone has the potential to cause problems: too few backchannels and a speaker appears unengaged, too many and they seem impatient. However, what is "too few" or "too many"? Maynard (1989, 1997) and Ward and Tsukahara (2000) claim that, even allowing for individual speaking styles, backchanneling is more frequent in Japanese than in English. There also appear to be differences not only in the frequency of responses but in what kind of cue can elicit backchannel responses. A phenomenon that typically elicits a response in one language does not necessarily do so in another language. For example, in studies of turn-taking cues in Dutch

(Caspers 2001) and in English (Wichmann and Caspers 2001), it was found that a contour that appears to cue backchannel in Dutch (a high-level tone) blocks backchannel in English. Such differences have implications for cross-cultural communication. A backchannel response elicited but not forthcoming, and also a response that is unsolicited and unexpected, can be perceived as "trouble" and interpreted negatively.

Attitude/interpersonal meaning

Brewer (1912) was not wrong in telling performers to establish sympathy with their audience, and the same is true for conversation. The expression of interpersonal meaning is crucially important to the success of communication. Mennen (2007) points out that the inappropriate use of intonation, and in particular its cumulative effect, can have negative consequences for the non-native speaker. Unlike segmental errors, suprasegmental errors are rarely recognized as such by native listeners, but simply misinterpreted as attitudes that the speaker did not intend. Pickering, Hu, and Baker (2012) claim rightly that "prosody contributes significantly to interactional competence and serves to establish a crucial collegial bond between speakers", and they conclude that "prosody in the English language classroom is key" (2012: 215). However, "attitude" remains the most elusive of meanings to capture analytically. What is it exactly about a speaker's "tone of voice" that can make an utterance sound "friendly" or "impolite"?

There are broadly two approaches to studying the correlates of perceived attitudes: the first is to look for features of an individual utterance that cause it to be perceived as "friendly, brusque, condescending", or any other of the many labels that can be used. The second is to focus on sequential relationships between utterances, and look for the meanings constructed by the similarity or differences between (usually consecutive) utterances rather than any features of an utterance itself. I will look at each approach in turn.

"Attitude" in utterances

Early work on English intonation, such as that of O'Connor and Arnold (1961), suggested that individual contours – falls, rises, fall-rises, and so on – carry independent meanings in conjunction with certain sentence types. However, intonation contours were ascribed so many "attitudinal" meanings that it became clear that the contour meant none of them. O'Connor himself noted that the topic of attitudinal intonation was "bedevilled by the lack of agreed categories and terms for dealing with attitudes" (1973: 270). A more abstract, reductive approach to the meaning of pitch contours is that of Cruttenden (1997), who sees falls and rises as "closed" and "open" contours, and Wichmann (2000), who refers to the same distinction in terms of "final" and "non-final". The rising tone of a yes-no

question is consistent with the "open" meaning of a rise, while the "closed" meaning of a falling nucleus is consistent with the syntactic completeness of a statement. This underlying meaning is used in Wichmann (2004) to explain why *please*-requests with low endpoints imply little optionality (the matter is final/closed) while a request ending high suggests that the matter is still open, giving the addressee greater optionality. Gussenhoven (2004), building on earlier work of Ohala (1994), has suggested that this distinction is ethological in origin, in other words it goes back to animal behavior, and that low pitch is associated with big (and therefore powerful) animals, while high pitch is associated with small and therefore less powerful animals. The big/small association has, he suggests, become encoded in prosody. But how does this relate to "attitude"?

I have argued in the past (e.g., Wichmann 2000) that some perceived (usually negative) attitudes arise simply because there is a mismatch between the hearer's expectations and what the speaker actually does. On the assumption that the speaker intends to convey something relevant to the conversation, the hearer will endeavor to infer what this meaning is. A *please*-request uttered with a falling contour assumes compliance, but the hearer may not feel that assumed compliance is appropriate and may infer something resembling an insult. Similarly, if an expression of gratitude such as *thank you* sounds casual when the hearer believes that greater gratitude is due, they will perceive the speaker as "rude" or "off-hand". While these choices may have been intentional, with the speaker aware of the implicature generated and prepared to deal with the consequences, it may also be an unintended mistake, which will disrupt the communication until the misunderstanding is resolved. In other words, if things go wrong, participants interpret prosodic "mistakes" as intentional messages and infer meaning accordingly.

Perceived "mismatches" – prosodic behavior that appears to diverge from the hearer's expectations, especially in cross-cultural situations – also arise in other areas of prosody. Some cultures, for example, tolerate silences between turns, while others value the apparent "enthusiasm" of overlapping speech. Cultural rules for turn-taking behavior are unconscious, and if they are broken, the participants assume that it reflects some intentional behaviour – reticence, aggressiveness, enthusiasm, and so on – rather than a simple error. Tannen (1981) notes the different attitude to turn-taking between New York Jewish speakers and non-New Yorkers. Overlap is "used cooperatively by New Yorkers, as a way of showing enthusiasm and interest, but it is interpreted by non-New Yorkers as just the opposite: evidence of lack of attention". In some cases, divergent behavior can be responsible for national stereotypes, such as "the silent Finn", because of the Finnish tolerance for long silences in conversation. Eades (2003) points to a problem arising from similar discrepancies in the interactional behavior between Australian English and Australian Aboriginal cultures. In Australian English interaction a long silence is unusual, and can cause discomfort, but Aborigines value silence, and do not regard it as a sign that the conversation is not going well (2003: 202–203). Eades is particularly concerned with the disadvantage for Aborigines in the context of

the courtroom, where silences "can easily be interpreted as evasion, ignorance, confusion, insolence, or even guilt" (2003: 203).

"Attitude" through sequentiality

The second, very different, approach to the prosodic expression of attitude has been suggested by research into prosodic entrainment or accommodation. The idea of "entrainment" goes back to observations in the seventeenth century of the behavior of pendulums, which gradually adapt to each other's rhythm. Conversation between adults frequently displays accommodation or convergence in both verbal and nonverbal behavior. Gestures, posture, and facial expressions can all mirror those of the interlocutor, while accommodation in speech includes changes to pronunciation and, at the prosodic level, pitch range, pausing, and speech rate. Whether this tendency to converge or accommodate is an automatic reflex or a socially motivated behavior is still a matter of debate (see the discussion in Wichmann 2012). There is no doubt an element of both, and the degree of accommodation may to some extent depend on the affinity felt between interlocutors (Nilsenova and Swerts 2012: 87). By mirroring the other's verbal and nonverbal signals it is possible to both reflect and to create a greater rapport with the other. Conversely, a failure to accommodate may reflect, or create, a distance between interlocutors.

We have already seen that this kind of rhythmic entrainment or adaptation may account for the timing of turns and backchannel responses. There is also evidence to suggest that a similar accommodation occurs in the choice of pitch "register". An early model of English intonation that contained an element of sequentiality is the discourse intonation model of David Brazil (e.g., Brazil, Coulthard, and Johns 1980), in particular his idea of "pitch concord", which involves matching pitch level across turns (see also Wichmann 2000: 141–142). An interactional account of pitch matching is also to be found in Couper-Kuhlen (1996), who suggests that when a speaker response echoes the previous utterance using the same register (i.e., relative to the speaker's own range), the response is perceived as compliant, whereas if it copies the pitch contour exactly it can be perceived as mimicry. A more recent longitudinal study by Roth and Tobin (2009) showed that prosodic accommodation between students and teachers correlated with lessons perceived as "harmonious".

It is this matching across turns, in addition to the phonological choices made within an utterance, which can generate – intentionally or unintentionally – a perceived "attitude". Conversational participants are expected to be "in time" and "in tune" with one another; failure to do so may suggest a lack of affinity, whether or not it was intended. The "attitude" that is then perceived by the hearer is a pragmatic inference that depends on the context of situation.

As Nilsenova and Swerts (2012) rightly point out, an awareness of accommodation behavior, and the signals it can send, may be important for learning situations. Above all, it reminds us that human communication does not consist of isolated utterances but that meaning is made jointly: as Tomasello puts it: "(h)uman communication is … a fundamentally cooperative enterprise" (2008: 6).

REFERENCES

Aijmer, K. 1996. *Conversational Routines in English: Convention and Creativity*, London: Longman.

Archer, D., Aijmer. K., and Wichmann, A. 2012. *Pragmatics. An Advanced Resource Book for Students*, London: Routledge.

Baumann, S. and Schumacher, P.B. 2011. (De-)accentuation and the processing of information status: evidence from event-related brain potentials. *Language and Speech* 55(3): 361–381.

Benus, S., Gravano, A., and Hirschberg, J. 2007. The prosody of backchannels in American English. In: *Proceedings ICPhS XVI Saarbrücken*, 1065–1068.

Brazil, D., Coulthard, M., and Johns, C. 1980. *Discourse Intonation and Language Teaching*, London: Longman.

Brewer, R.F. 1912 Speech. In: *Voice Speech and Gesture. A pPractical Handbook to the Elocutionary Art*, R.D. Blackburn (ed.), Edinburgh: John Grant.

Brown, G., Currie, K., and Kenworthy, J. 1980. *Questions of Intonation*, Baltimore: University Park Press.

Caspers, J. 2001. Testing the perceptual relevance of syntactic completion and melodic configuration for turn-taking in Dutch. In: *Proceedings of Eurospeech*.

Chafe, W. 1979. The flow of thought and the flow of language. *Syntax and Semantics, vol. 12: Discourse and Syntax*, Academic Press Inc.

Chen, A., Den Os, E., and De Ruiter, J.P. 2007. Pitch accent type matters for online processing of information status. *The Linguistic Review* 24(2): 317–344.

Couper-Kuhlen, E. 1996. The prosody of repetition: on quoting and mimicry. In: *Prosody in Conversation*, E. Couper-Kuhlen and M. Selting (eds.), Cambridge: Cambridge University Press.

Couper-Kuhlen, E. and Ford, C.E. 2004. *Sound Patterns in Interaction*, Amsterdam: John Benjamins.

Cruttenden, A. 1997. *Intonation*, 2nd edition, Cambridge: Cambridge University Press.

Dahan, D., Tanenhaus, M.K., and Chambers, C.G. 2002. Accent and reference resolution in spoken language comprehension. *Journal of Memory and Language* 47: 292–314.

Deterding, D. 2012. Issues in the acoustic measurement of rhythm. In: *Pragmatics and Prosody in English Language Teaching*, J. Romero-Trillo (ed.), 9–24, Dordrecht: Springer.

Eades, D. 2003. The politics of misunderstanding in the legalsystem: Aboriginal speakers in Queensland. In: *Misunderstandings in Social Life*, J. House, G. Kasper, and S. Ross (eds.), 199–226, London: Longman.

Edelsky, C. 1981. Who's got the floor? *Language and Society* 10: 383–421.

Fry, D.B. 1955. Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America* 27(4): 765–768.

Fry, D.B. 1958. Experiments in the perception of stress. *Language and Speech* 1: 126–152.

Gordon, M. and Applebaum, A. 2010. Acoustic correlates of stress in Turkish Kabardian. *JIPA* 40(1): 35–58

Gussenhoven, C. 2004. *The Phonology of Tone and Intonation*, Cambridge: Cambridge University Press.

Halliday, M.A.K. and Hasan, R. 1976. *Cohesion in English*, London: Longman.

Heldner, M. 2011. Detection thresholds for gaps, overlaps, and no-gap-no-overlaps. *Journal of the Acoustical Society of America* 130(1): 508–513.

Hirschberg, J. and Ward, N. 1995. The interpretation of the high-rise question contour in English. *Journal of Pragmatics* 24: 407–412.

Jurafsky, D., Bates, R., Coccaro, N., Martin, R., Meteer, M., Ries, K., Shriberg, E., Stolcke, A., Taylor, P., and Van Ess-Dykema, C. 1997. Automatic detection of discourse structure for speech recognition and understanding. In: *Proceedings of the 1997 IEEE Workshop on Automatic Speech Recognition and Understanding*.

Kong, E. 2004. The role of pitch range variation in the discourse structure and intonation structure of Korean. In: *Proceedings of Interspeech*, 3017–3020.

Lehiste, I. 1979. Perception of sentence and paragraph boundaries. In: *Frontiers of Speech Research*, B. Lindblom and S. Ohman (eds.), London: Academic Press.

Levis, J. and Pickering, L. 2004. Teaching intonation in discourse using speech visualisation technology. *System* 32: 505–524.

Local, J. 1992. Continuing and restarting. In: *The Contextualization of Language*, P. Auer and A. di Luzio (eds.), 273–296, Amsterdam: Benjamins.

Local, J. and Walker, G. 2004. Abrupt joins as a resource for the production of multi-unit, multi-action turns. *Journal of Pragmatics* 36(8): 1375–1403.

Maynard, S.K. 1989. *Japanese Conversation*. Norwood, NJ: Ablex.

Maynard, S.K. 1997. Analyzing interactional management in native/non-native English conversation: a case of listener response. *International Review of Applied Linguistics in Language Teaching* 35: 37–60.

Mennen, I. 2007. Phonological and phonetic influences in non-native intonation. In: *Non-native Prosody: Phonetic Descriptions and Teaching Practice*, J. Trouvain and U. Gut (eds.), 53–76, The Hague: Mouton de Gruyter.

Nakajima, S. and Allen, F.A. 1993. A study on prosody and discourse structure in cooperative dialogues. *Phonetica* 50: 197–210.

Nilsenova, M. and Swerts, M. 2012. Prosodic adaptation in language learning. In: *Pragmatics and Prosody in English Language Teaching*, J. Romero-Trillo (ed.), 77–94, Dordrecht: Springer.

O'Connor, J.D. 1973. *Phonetics*, Harmondsworth: Penguin Books.

O'Connor, J.D. and Arnold, G.F. 1961. *Intonation of Colloquial English*, London: Longman.

Ohala, J.J. 1994. The frequency code underlies the sound-symbolic use of voice pitch. In: *Sound Symbolism*, L. Hinton, J. Nichols, and J.J. Ohala (eds.), 325–347. Cambridge, Cambridge University Press.

Peppé, S. and McCann, J. (2003). Assessing intonation and prosody in children with atypical language development: the PEPS-C test and the revised version. *Clinical Linguistics and Phonetics* 17(4/5): 345–354.

Pickering, L., Hu, G., and Baker, A. 2012. The pragmatic function of intonation: cueing agreement and disagreement in spoken English discourse. In: *Pragmatics and Prosody in English Language Teaching*, J. Romero-Trillo (ed.), 199–218, Dordrecht: Springer.

Pierrehumbert, J.B. 1987. *The Phonology and Phonetics of English Intonation*. PhD thesis 1980, Indiana University Linguistics Club.

Pierrehumbert, J.B. and Hirschberg, J. 1990. The meaning of intonation contours in the interpretation of discourse. In: *Plans and Intentions in Communication and Discourse*, P.R. Cohen, J. Morgan, and M.E. Pollack (eds.), 271–311, Cambridge, MA: MIT Press.

Piske, T. 2012. Factors affecting the perception and production of L2 prosody: research results and their implications for the teaching of foreign languages. In: *Pragmatics and Prosody in English Language Teaching*, J. Romero-Trillo (ed.), 41–59, Dordrecht: Springer.

Riesco, B.S. and Romero-Trillo, J. 2008. The acoustics of 'newness' and its pragmatic implications in classroom discourse. *Journal of Pragmatics* 40 (6) 1103–1116.

Roth, W.-M. and Tobin, K. 2009. Solidarity and conflict: aligned and misaligned prosody as a transactional resource in intra- and intercultural communication involving power differences. *Cultural Studies of Science Education*. doi: 10. 1007/s11422-009-9203-8.

Sacks, H., Schegloff, E.A., and Jefferson, G. 1974. A simplest systematics for the organisation of turn-taking in conversation. *Language* 50: 696–735.

Shriberg, E., Bates, R., Stolcke, A., Taylor, P., Juraafsky, D., Ries, K., Coccaro, N., Martin, R., Meteer, M., and van Ess-Dykema, C. 1998 Can prosody aid the automatic classification of dialog acts in conversational speech? *Language and Speech* 41 (3–4) 443–492.

Sluijter, A.M.C. and Terken, J.M.B. 1993. Beyond sentence prosody: paragraph intonation in Dutch. *Phonetica* 50: 180–188.

Szczepek Reed, B. 2011. Beyond the particular: prosody and the coordination of actions. *Language and Speech* 55(1): 13–34.

Tannen, D. 1981. New York Jewish conversational style. *International Journal of the Sociology of Language* 30: 133–149.

Thompson, S.E. 2003. Text-structuring metadiscourse, intonation and the signalling of organisation in academic lectures. *Journal of English for Academic Purposes* 2: 5–20.

Tomasello, M. 2008. *Origins of Human Communication*, London: MIT Press.

Ward, N. and Tsukahara, W. 2000. Prosodic features which cue back-channel responses in English and Japanese. *Journal of Pragmatics* 32: 1177–1207.

Wells, J. 2006. *English Intonation*, Cambridge: Cambridge University Press.

Wichmann, A. 2000. *Intonation in Text and Discourse*, London: Longman.

Wichmann, A. 2004. The intonation of *please*-requests: a corpus-based study. *Journal of Pragmatics* 36: 1521–1549.

Wichmann, A. 2005. *Please* – from courtesy to appeal: the role of intonation in the expression of attitudinal meaning. *English Language and Linguistics* 9(2): 229–253.

Wichmann, A. and Caspers, J. 2001. Melodic cues to turntaking in English: evidence from perception. In: *Proceedings, 2nd SIGdial Workshop on Discourse and Dialogue*, Aalborg, Denmark.

Wilson, M. and Wilson, T.P. 2005. An oscillator model of the timing of turn-taking. *Psychonomic Bulletin and Review* 12(6): 957–968.

Yngve, V. 1970. On getting a word in edgewise. In: *Papers from the Sixth Regional Meeting of the Chicago Linguistic Society*, 567–577.

Yule, G. 1980. Speakers' topics and major paratones. *Lingua* 52: 33–47.

11 Pronunciation and the Analysis of Discourse

BEATRICE SZCZEPEK REED

Introduction

Spoken interaction relies entirely on the way in which utterances are physically delivered. While the pronunciation of vowels and consonants can tell us a lot about the identity of a speaker in terms of, for example, where they come from, their speech melody, rhythm, and tempo will help create specific discourse meanings uniquely fitted to a given conversational moment. Producing vowels and consonants involves what phoneticians call *articulation*, that is, the pronunciation of individual speech sounds. Sounds are conceived of as *segments* of words and are therefore often referred to as representing the *segmental* level of speech. Features such as rhythm, intonation and tempo, on the other hand, are frequently referred to as *suprasegmentals*, as they apply not to individual sounds, but to entire words, or even utterances: they occur above the level of the single segment. For the analysis of spoken interaction the suprasegmental level of talk is the most relevant, as speakers employ it to subtly manipulate the pragmatic meaning of their utterances. Therefore, this chapter is primarily concerned with the suprasegmental aspects of speech.

Another term that is frequently used for suprasegmentals is *prosody*, often defined as the musical aspects of speech: pitch, loudness, and time. In the following section the role of prosodic features for the accomplishment of conversational actions will be considered, and it will be discussed whether it is possible to assign specific discourse functions to individual features. Subsequently, issues surrounding the learning and teaching of pronunciation will be presented, and the argument will be made that in order to achieve interaction successfully and fluently in a second language, it is not necessary to speak with "native-like" prosody.

The Handbook of English Pronunciation, First Edition. Edited by Marnie Reed and John M. Levis.

The role of prosody for discourse

Research on prosody in conversation has shown that the pitch, loudness, and timing of utterances play a vital role in shaping the social actions that speakers perform through language. However, the fact that speakers do not follow a pre-scripted plan but instead continuously create new interactional situations, with new contingencies and risks, means that the role of prosody is a complex one. Nevertheless, there are some contexts in which certain prosodic features seem to be used regularly and systematically. Below we consider conversational turn-taking, sequence organization, and individual actions, such as repair and reported speech.

The examples of naturally occurring talk presented in this chapter are transcribed according to an adapted version of the GAT conventions (Selting et al., 1998), which can be found in the Appendix. Briefly, punctuation marks are used to denote phrase-final pitch movements, such as commas for rise-to-mid and periods for fall-to-low, and capital letters are used to denote levels of stress. The rationale for using such a system, rather than IPA transcription, for example, is to allow the analyst to incorporate prosodic (rather than phonetic) information while still providing an accessible transcript to a broad readership.

Turn-taking

One of the most important conversational activities is *turn-taking*, that is, speakers' moment-by-moment negotiation over who speaks next, and for how long. Here, prosody is used as an important cue for whether an utterance, or turn, is potentially complete, or whether its speaker intends to continue talking.

In the following example, Rich is telling his brother Fred about life without a girlfriend.[1] In theory, Fred could come in to speak after line 2 or line 4; however, the intonation at the end of those turns is level, as indicated by the dash symbol in the transcript. Fred only starts speaking when Rich has produced low falling intonation at the end of his turn at lines 5–6, indicated by a period.

1. SBC047 On the Lot

```
1   Rich:  it's LONEly coming home after pUtting in t- twelve hours
2          on the LOT - =
3          and wOrking All DAY and;
4          yOU know wOrking all EVEning - =
5          and then you don't have Any(.)bOdy to come hOme and
6          SHARE it with.
7   (0.32)
8   Fred:  YEAH;
9   (0.54)
10         .hh a- are y- are yOU WORKing twelve hours?
```

Another piece of evidence that Rich has finished talking after line 6 is the pause at line 7: he does not say any more after he has produced the low falling pitch movement. This example demonstrates a regular occurrence for British and American standard varieties of English, where potential next speakers often wait until a current speaker has produced a low falling intonation contour before they come in to speak next. Of course, intonation is not the only factor affecting turn-taking decisions. Firstly, there are other prosodic features that play a role. Speakers usually slow down slightly towards the end of their turn and tend to lengthen the final syllable; their speech also decreases in loudness; and in some cases the last syllable takes on creaky voice quality. Secondly, nonprosodic features play an important role. Ford and Thompson (1996) show that it is a combination of grammatical, pragmatic, and prosodic cues that allows conversational participants to judge whether a speaker is finished or not, that is, a speaker has typically finished a sentence and the overall point they are making in terms of content before others come in to speak.

While the above example is representative of standard varieties of British and American English, the prosodic cues for turn-taking vary considerably across accents and dialects. For example, in Tyneside English, spoken in the North East of England, the prosody for turn completion is either a rise or a fall in pitch on the last stressed syllable, combined with a slowing down towards the end of the turn, a sudden increase and decrease in loudness on the last stressed syllable, and lengthening of that syllable (Local, Kelly, and Wells 1986). Similarly, the prosody of turn completion in London Jamaican (Local, Wells, and Sebba 1985) and Ulster English (Wells and Peppe 1996) varies from standard varieties of English.

While turn transition after low falling pitch is a frequently occurring phenomenon, it would be wrong to assume that every time a speaker uses a low fall in pitch they automatically stop speaking and another participant comes in. While discourse participants orient to systematic uses of conversational resources, they nevertheless negotiate each social action individually. This is also true for turn-taking, which means that at each potential turn completion point current speakers may choose to continue or not; and next speakers may choose to come in or not. The systematics for turn-taking have been described in a seminal paper by Sacks, Schegloff, and Jefferson (1974). The following example shows this clearly. At lines 5 and 10, Michael produces potential turn completion points, at which his intonation falls to low. Both are followed by pauses, showing that Michael is leaving the floor to be taken up by his co-participants. When this does not happen, he himself continues speaking.

2. SBC017 Wonderful Abstract Notions

```
1 Michael: but there's ONE techNOLogy that's uh:m;
2 (0.19)
3          gonna overtake THA:T and that's;
4 (0.17)
```

```
5           DNA research.
6  (0.12)
7           WHICH is LIKE (0.11) a TOtal SCAM at thIs point still
8           it's they're just like (0.18) bomBARDing;
9  (0.75)
10          .h ORganisms with radiAtion to see what comes UP.
11 (0.31)
12          .hh you KNOW;
13          we have vEry little conTROL over it;
14          but once we ↑DO;
15 (0.58)
16          .hh we'll be able to prOgrA:m biOlogy as WELL.
17 (0.83)
18 Jim:     well THA:T'S pretty frIghtening cOncept.
19 Michael: it IS frIghtening but-
20 (0.3)
21          [uhm
22 Jim:     [we cAn't even control our FREEways.
```

It is only at line 18, and after a considerable pause following another low falling turn completion point that Jim comes in to speak. His utterance ends in low falling intonation, and is immediately responded to by Michael. However, at line 22 we see another local variation of the turn-taking system: Jim comes in to speak even though Michael's previous turn (line 19) is neither grammatically nor prosodically complete.

The above example demonstrates that we cannot assume a straightforward form-function relationship between prosodic features and discourse actions. Speakers may routinely orient to certain patterns, but nevertheless negotiate individual sequences afresh. Furthermore, in the same way that we cannot assume that speakers always implement turn-taking after prosodic turn-taking cues, we also cannot assume that the prosodic cues for turn-taking are always the same. While there is a strong orientation to low falling intonation, many other patterns may appear at the end of turns depending on the immediate context (Szczepek Reed 2004). In the following example, Joanne describes her favorite holiday destination, Mexico, by listing the many things she likes about it.

3. SBC015 Deadly Diseases

```
1 Joanne: BEAUtiful BEAUtiful blue hehe blue WAter,
2         and and .hh WARM Water -
3         and like CORal and TROPical F:ISH - =
4         and inCREDible r- like reSORT -
5 (1.26)
6         lIke uh::m;
7         <<p> hoTEL:S,
```

```
8          and REStaurants,>
9  Ken:    .hh Oh when wE were there LAST;
10         we- th- it was JUST after an eLECtion;
```

The intonation for each list item is either slightly rising (lines 1, 7, 8) or level (lines 2, 3, 4). Neither pitch movement is routinely employed as a turn completion cue; however, in this instance, another prosodic feature plays an important role. Towards the end of her list, Joanne's voice becomes softer (lines 7–8), indicated by <<p>> for "piano". As the turn fades out, Ken comes in to speak (line 9) after a slightly rising intonation contour. There is no further talk from Joanne, which suggests that she indeed had not planned to continue speaking. It is also relevant that her earlier pause of 1.26 seconds (line 5) and her use of the tokens *like uhm* (line 6) indicate local difficulties in the construction of the turn, while the rising pitch on the final two list items projects the *potential* for more items, rather than necessarily their upcoming delivery.

Prosody also plays a role when turn-taking becomes problematic. French and Local (1986) describe how it is primarily through prosody that participants show whether they consider themselves to be the rightful turn holder, in which case they increase their loudness in the face of an interruption, or whether they are illegitimately interrupting, in which case they increase both loudness and pitch register. Participants who are being interrupted typically raise their overall loudness until the interrupter drops out. See, for example, the following excerpt, in which Angela interrupts Doris at line 4.

4. SBC011 This Retirement Bit

```
1 Doris:    I'm not a very good PILL taker.=
2           I'm re-
3           i THINK i'm [reSENTing;
4 Angela:               [I'm not EIther but [i get-
5 Doris:                                    [<<f> I'm resenting> this
6           MEDicine.
7           and I think it's conTRIButing to my PROBlems.
8           i REALly DO.
```

In response to Angela's interruption at line 4, Doris increases her loudness (lines 5–6), indicated in the transcript by <<*f*>> for *forte*. She does so only for a very short part of her utterance (*I'm resenting*), until Angela has stopped speaking, after which Doris returns to her default loudness.

In the following example, mother Patty and daughter Steph are discussing Steph's SAT scores with Steph's friend Erika.

5. SBC035 Hold my Breath

```
1 Steph:  i KNOW what the trIcks are.=
2         that's ALL you need to KNOW.
```

```
3 Erika:   TEACH them to [me.
4 Steph:                 [<<f> the Only [way you can-
5 Patty:                                [<<f+h> but whAt you HAVE to
6          remEm[ber I:s that-
7 Steph:        [<<f+h> the Only way you can SCORE high> <<dim> is
8          if you READ a lot.>
9          [THAT'S ALL.
10 Patty: [<<f+h> what you HAVE to re[MEMber is;>
11 Steph:                           [you CAN'T study;
12 Patty: <<f> that the SAT> is nOt a whole mEAsure of who you
13         ARE.>
```

Patty interrupts Steph at line 5, at a point where Steph has clearly not finished speaking. Patty does so with high overall pitch register and high overall loudness. At line 7, Steph also raises her loudness and pitch register, but reduces both once her mother has dropped out. Patty interrupts again, with high loudness and overall pitch, but also returns, first, to her default pitch register, and then to her default loudness as Steph drops out (line 12).

In considering these examples we must bear in mind that increased loudness and high pitch register may accomplish many other things besides interruptions in conversation and that interruptions may not always display these features, depending on the type of interruption a speaker is engaged in. While it is the case that participants in conversation use prosodic features systematically, they also do so flexibly, as each instance emerges as part of its specific interactional context.

Sequence organization

Another primary action participants are involved in during spoken interaction is *sequence organization* (Schegloff 2007). This term refers to the way in which speakers organize larger conversational projects, such as narratives, complaints, or requests. Here prosody also plays an important role. For example, Couper-Kuhlen (2004) suggests that when speakers begin a new sequence in conversation, they usually do so by stepping up to a higher pitch. Similarly, Local (1992) shows that when speakers design an utterance that was interrupted as a "restart", they do so with a change to higher pitch, whereas when they design it as a "continuation" of prior talk they do so at the same pitch level as the prior utterance. However, Szczepek Reed (2006, 2009, 2012a) has shown that it is not so much a specific prosodic pattern, or even feature, that is relevant for designing talk as continuing a previous sequence or starting a new one. What seems more relevant is whether participants repeat the prior speaker's overall prosodic design, or not.

In the following excerpt, two short sequences are accomplished with prosodic repetition, or "matching" (Szczepek Reed 2006). At line 3, Alan initiates repair on Jess's previous turn, that is, he indicates that he has a problem with it: Jess claims that a book she has been looking for is not held by the British Library, which Alan responds to with *what?* He does so with a high pitch register (line 3).

6. BSR REC6

```
1  Jess:   the british LIbrary doesn't even hAve it though.
2  (0.4)
3  Alan:   <<h> WHAT - >
4  Jess:   <<h> YEAH - >
5  (0.4)
6          because like an amErican -
7  Alan:   OH yeah;
8  Jess:   BOOK;
9  Alan:   they USED to have every (.) english bOOk of course dOn't
10         they;
11 Jess:   oh IS it;
12 (0.3)
13 Alan:   YEAH::;
14 Jess:   i had ↑QUITE a [lot of Other ones tOO;
15 Alan:                  [i think it might be the LAW (.) they have
16         to give them all-
17 Jess:   <<h> oh REALly;>
18 Alan:   <<h> YEAH:: i thInk> [so:;
19 Jess:                       [SHI::T;
```

In response to Alan's repair initiation, Jess provides the repair, that is, she confirms that what Alan had found hard to believe is indeed true (line 4). What is interesting is that while Jess's pitch at line 1 is in her default range, her pitch at line 4 matches Alan's: his repair initiation is produced with high pitch register and so is Jess's repair. The pitch matching can be seen in Figure 11.1 (Alan's turn is represented in the top tier, Jess's turn in the bottom tier). Shortly afterwards at line 17, Jess issues a news receipt, *oh really*, of Alan's previous informing at lines 15–16. The news receipt is again produced with a high overall pitch register and so is Alan's response, *yeah I think so* (line 18). Once again, a response is designed as matching the prosodic design of the turn it is designed to respond to.

Both sequences in the above excerpt are adjacency pairs (Schegloff 2007), that is, they are sequences in which one turn (a so-called First Pair Part) initiates and makes relevant a certain type of action by another speaker (a Second Pair Part). Typical adjacency pairs are question-answer, greeting-return greeting, or, as in this case, repair initiation-repair and news receipt-confirmation. The matching of the prosodic design plays an important role for the second turn to be heard and treated as a Second Pair Part. If second speakers do not match first speakers' prosody, their responses may not be treated as appropriate. See, for example, the next excerpt, in which Julie greets Tricia, a 9-year-old child. Julie's first greeting is delivered with a high pitch register and a musical interval. Tricia's next turn is produced with a low pitch register.

7. BSR Farm (no recording)

```
1 Julie:   <<h + musical interval> HI TRIcia - >
2 Lisa:    <<l> hellO;>
```

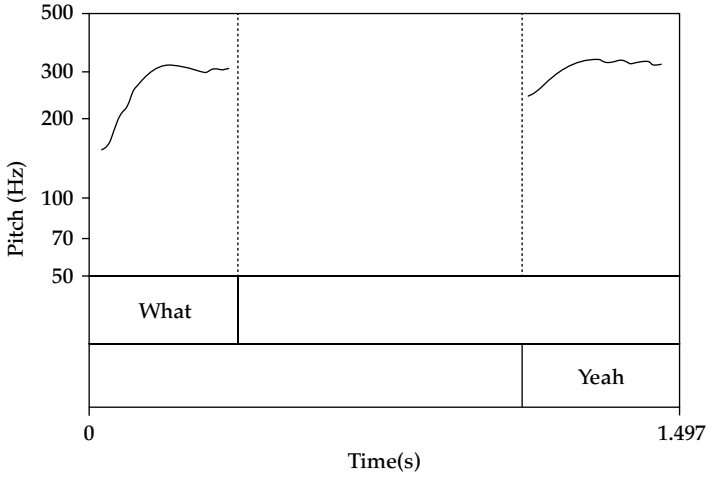


Figure 11.1 Pitch matching in excerpt (4), lines 3–4.

```
3 Julie:   <<l> hellO;>
4 Mum:     <<h > HI TRISH -
5          yOU alRIGHT,>
6 Lisa:    <<l> NO;>
```

What is noticeable about this excerpt is that Julie issues another greeting at line 3. This shows that she does not treat Tricia's low pitched *hello* as a return greeting to her own earlier high pitched greeting; instead, she treats it as a new first greeting, requiring a response. Her return greeting matches Tricia's low pitch.

This example shows that with regard to the starting of new sequences in conversation, it is not necessarily the precise nature of the prosodic design that is relevant (i.e., high pitch), but instead the presence or absence of prosodic matching more generally. Turns that match the prosody of prior turns may be treated as responding. Turns that do not match the prosody of previous talk may be treated as starting something new – even if the speaker uses low pitch, as in excerpt (7).

Conversational actions

Besides turn-taking and sequence organization, which have to be achieved throughout speakers also make use of prosody in their accomplishment of individual social actions. In most cases, prosody is not the only feature that implements actions, but there are some instances in which it plays a primary role. For example, Selting (1996) shows that the German repair initiation "bitte" ("pardon") is used in two different prosodic variants, which are treated by

recipients as implementing two different conversational actions. Both versions of "bitte" are produced with rising intonation, but in one case the pitch rises considerably higher than in the other and the overall loudness is increased. Selting shows that while "bitte" with a default pitch span and loudness is treated as initiating repair over mechanical issues, such as an acoustic issue or other understanding problem, "bitte" with a wide pitch span and increased loudness is treated as a cue for astonishment. In the second case, next speakers do not repeat what they said, as they do after "bitte" with default prosody. Instead, they display accountability, thus showing that they heard the loud and high-pitched "bitte" as initiating repair over the content of their previous turn, rather than over acoustic features.

Another activity for which prosody seems to be crucial is participants' quoting of others. Couper-Kuhlen (1996) shows that there is a clear interactional distinction between simply repeating what another speaker said and mimicking it. This distinction is achieved primarily through prosody, particularly pitch register. Couper-Kuhlen describes two variations of male speakers' repeating female speakers' talk. In both cases, the men match the women's pitch register. However, they do so either on a relative or on an absolute scale. If pitch register is matched on a relative scale that means that a male speaker repeats the word or phrase of an immediately prior female speaker in the same pitch register, but relative to his own voice range. Thus, if a female speaker is speaking in an upper-mid register, the male speaker will repeat the female speaker's words in what is an upper-mid register for him. This is the default case, and is treated as unmarked repetition by participants themselves. If, on the other hand, the quoting male speaker matches the female pitch register on an absolute scale, this means he uses exactly the same pitch as the woman, thus speaking extremely high in his own voice range. This is treated by participants as mimicry and a form of implicit criticism.

Klewitz and Couper-Kuhlen (1999) consider quoting nonpresent speakers, and compare prosody to the use of quotation marks in written texts. They show that while a change to a different prosodic pattern may indicate the onset of reported speech, spoken discourse is much more flexible than written punctuation and does not require prosodic marking to continue for the whole stretch of reported speech. A change to a high pitch register, for example, may be enough to indicate that reporting has begun, even though it may not be sustained throughout the entire turn. Interestingly, speakers may also project upcoming reported speech by adopting the prosodic design before the actual reported speech sequence has begun, as in the following excerpt.

8. SBC006 Cuz

```
1 Alina: (JOY) talked the whole time;=
2        <<falsetto+extra high+all> in a voice like THIS -
3 (0.44)
4        <<higher falsetto> HI:: ((alina)) -
```

```
5          i'm so HAPpy to see YOU::;>>
6          <<laughing> and we're going - >
7  (0.4)
8          .hh <<h> GO::D;
9  (0.34)
10         turn the VOLume <<laughing> dOwn;>
```

In this excerpt, Alina voices the speech of a nonpresent person referred to here as Joy. She uses an extreme prosodic format, involving falsetto voice quality in combination with extremely high pitch register and fast speech rate. However, she starts using these features already on her pre-quotation talk at line 2 (*in a voice like this*), thus indexing the *voice* before actually voicing it. Following the reported speech sequence, Alina returns to her default voice quality, pitch register, and speech rate (line 6).

Summary

In this section the role of prosody for conversation has been outlined, with a focus on the two main discursive activities that speakers are involved in almost continuously in interaction: turn-taking and sequence organization. While it is clear that prosodic features are important in speakers' negotiation over these activities, it is not at all easy to establish a specific form-function relation for any given prosodic feature. For example, low falling intonation may at times be interpreted as a cue for turn completion; at other times, it may not be. Similarly, while slightly rising intonation can be a cue for a speaker's intention to continue talking, at times co-participants may come in after a slightly rising contour without being treated as illegitimately interrupting. This points to the multilayered role that prosody plays in conversation: while it might be treated as a turn-taking cue in some instances, which might require falling pitch, in others its main role may be to contribute to an utterance as a list item, which may require rising pitch. Prosody also contributes to linguistic distinctions. For example, increases in pitch, loudness, and lengthening cause a syllable to be perceived as stressed, while intonation helps listeners separate syntactic phrases. Furthermore, speakers use prosody as a cue for displaying affect and stance (Reber 2012). On the other hand, most actions in conversation are not accomplished only through prosody, but through other interactional resources, such as grammar, word choice, gaze, and gesture. It may be that at those times when prosody is not employed as a turn-taking cue other resources are used in its place.

However, while individual pronunciation features cannot easily be assigned specific discourse functions, there are broader interactional activities, such as ending an activity or starting a new one, which are systematically accomplished prosodically. As the examples above show, speakers orient to a distinction between repeating and not repeating a prior prosodic design and treat it as a distinction between continuing an ongoing sequence and a new beginning.

For the analysis of discourse it is vital to maintain a flexible perspective on prosody that allows for an understanding of interaction as emerging and locally negotiated. For the teaching and learning of prosodic pronunciation features such

a flexible perspective presents a potential problem, as it is much easier to learn and teach specific functions of prosody than to acquire pronunciation as a resource for locally accomplished actions. In the following section, these issues are considered in more detail.

Implications for learning and teaching pronunciation

Since it is impossible to speak without prosody – speech will almost always be produced at some pitch level, with some form of intonation and loudness, with some form of voice quality, etc. – the question arises as to which aspects of prosody differ across languages. One might argue that it is only those features that differ between a learner's L1 and L2 that should be taught in the language classroom. However, the discursive perspective detailed above suggests that what counts in interaction is not necessarily the "correct" pronunciation of utterances according to "native" speaker standards, but the appropriate use of prosodic features in any given context of social interaction. Thus, a different argument might be that if a language learner is able to use prosody in a way that implements social actions appropriately, then the influence of their L2 phonology does not matter interactionally outside those action contexts.

Jenkins (2000) argues that the primary goal for English pronunciation teaching should be intelligibility and, for learners who use English mainly as a lingua franca, international intelligibility. That is, only those pronunciation features should be taught that contribute to internationally intelligible speech, a suggestion that has inspired much debate (Levis 2005; Dziubalsak-Kołaczyk and Przedlacka 2005). Jenkins makes her argument against the background of academic discussions of English as a global, rather than a regional language, and the consequences this has for language learning and teaching. It is possible to develop this argument further, and take not only intelligibility but the successful accomplishment of actions in interaction as the criterion for teaching and learning pronunciation features. Regarding prosody, this argument is a particularly powerful one, given the flexible use of prosodic parameters by "native" speakers compared to segmental pronunciation features.

In the following we explore these issues by looking at speech rhythm, a prosodic feature whose form varies widely across languages. While features such as loudness, voice quality, and even pitch may have certain universal applications due to their close relation to physical sound production, time-related features such as syllable lengthening, stress, and rhythm have closer connections to the linguistic structures of each language.

Speech rhythm: stress timing and syllable timing

Rhythm is a feature of all languages, as all speech adheres to some form of regularity in its fluent organization of words and syllables. However, describing languages as rhythmic does not mean that speech in those languages is perfectly

isochronous, that is, absolutely regular. Rhythm is very much a perceptive phenomenon, and listeners will hear regularity even if the placement of rhythmic beats deviates to some extent from perfect isochrony. Nevertheless, most speech shows some form of regularity, even if languages differ greatly in their rhythmic organization. Phoneticians typically identify languages as belonging to one of two "rhythm classes": stress timing and syllable timing (Pike 1945; Abercrombie 1967), with most languages located somewhere along this spectrum of extremes (Dauer 1983; Miller 1984). Standard British English is classed as a highly stress-timed variety.

In short, stress timing refers to the perception of stressed syllables as being placed on rhythmic beats, as in:

a total scam at this point (excerpt 2, line 7)

This may not seem particularly remarkable, but becomes more so in utterances where stressed syllables are separated by unequal numbers of unstressed syllables, as in:

not a whole measure of who you are (excerpt 5, lines 12-13)

In order for the stressed syllables in this last utterance to be perceived as a rhythmic pattern, the unstressed syllables between them must be spoken in a similar time interval, even though there are only two unstressed syllables following the first "beat" (*a whole*) but four (*-sure of who you*) following the second. As a result, the unstressed syllables following the second beat must be produced more quickly and will therefore be much shorter in duration than those following the first beat. This explains why stress timing is determined by measuring how much syllable duration varies in a given language or variety: languages in which syllable duration varies a lot are typically classed as tending towards stress timing; languages in which syllable duration is more equal are classed as tending towards syllable timing, which involves a perception of *each syllable* as a rhythmic beat in itself. As a result, syllable-timed speech is sometimes described as having more of a "staccato" rhythm (Brown 1988).

Speech rhythm in conversation

A common perception of English rhythm is that content words are stressed and function words are unstressed. However, as the examples above demonstrate this is not an accurate description of natural talk. In real-life conversation participants frequently stress function words, depending on the pragmatic meaning they are conveying and the social action they are engaged in.

In British English conversation, speech rhythm has been found to play an important role for turn-taking. Auer, Couper-Kuhlen, and Müller (1999) show that in British English, next speakers integrate their talk rhythmically into the rhythm of a previous speaker. They do so by placing their first stressed syllable

on what has been projected by the previous speaker's turn as the next rhythmic beat. Auer et al. show that this type of rhythmic integration is the default case for British English conversation, whereas producing a turn too early, or too late, with respect to a previous rhythm is treated by participants as a cue for conversational trouble. In the following example form a radio phone-in programme, the radio host's greeting is delivered with a clear rhythmic pattern (line 2). In his reply, the caller places is return the greeting token *hi* precisely on the next rhythmic beat (line 4).

9. BE Scientist: Roger

```
1 Host:       joining us on the show;
2             ROger's in CLACton hi ROger.
3 (0.16)
4 Caller:     HI.
5 Host:       now we've gOt uh just a MINute or two LEFT;
```

Figure 11.2 shows how the caller's production of *hi* overlaps almost perfectly with the onset of the projected next beat. The vertical dotted lines indicate rhythmic beats, while the bold vertical lines in the text tier indicate the onset of vowels in stressed syllables. In speech rhythm research it is customary to measure rhythmic beats from the vowel onset, rather than the onset of the syllable, due to

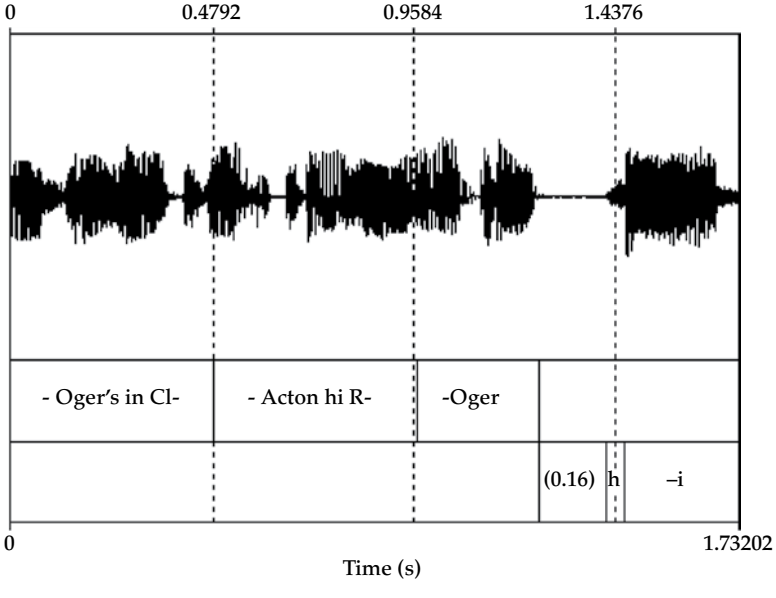


Figure 11.2 Rhythmic integration (Szczepek Reed 2009: 1234). Reproduced by permission of Elsevier.

the wide variations of syllable onsets in English.[2] The first interval was measured (0.4792 s) and then automatically superimposed over the rest of the waveform. The host's turn is represented in the top tier in the figure, the caller's turn in the bottom tier.

By integrating their next turn rhythmically into a prior speaker's turn, participants also achieve *interactional* integration and conversational fluency without noticeable gaps or overlaps (McCarthy 2009). Thus, regarding the learning and teaching of English speech rhythm, a vital question to ask is whether learners of English are able to accomplish integrated turn beginnings or if the rhythm of their first language impacts on their pronunciation to the effect that turn-taking is impeded.

Speech rhythm in conversations between syllable-timed and stress-timed speakers of English

Speech rhythm seems to affect the pronunciation of learners of English considerably, particularly if the speaker's first language has a tendency towards syllable timing (Adams 1979; Anderson-Hsieh, Johnson, and Koehler 1992; Anderson-Hsieh and Venkatagiri 1994; Bond and Fokes 1985; Brown 1988; Low 2006; Taylor 1981). The main influence of syllable-timed rhythm is on the pronunciation of unstressed syllables, such as weak forms. Learners of English whose first language has a tendency towards syllable timing may produce both stressed and unstressed syllables with relatively equal duration, thus making it difficult for listeners from a stress-timed background to identify which syllables are being stressed and which are not. From a conversational perspective, the main question is whether speakers of English with syllable-timed rhythm accomplish turn-taking successfully given the important role speech rhythm plays for the organization of speaker change. In a study of interactions between speakers of British English (BE) and Singapore English (SE), Szczepek Reed (2010, 2012b) investigated the rhythm and timing of turn transitions. The first language of the Singapore English speakers was Mandarin and all SE speakers had learned English from the age of 6. Both Mandarin and Singapore English have been classified as syllable-timed languages (Benton et al. 2007; Chen et al. 2001; Deterding 2001; Low, Grabe, and Nolan 2000). The study found that in any given conversation between a third and half of all turn transitions from the BE speaker to the SE speakers were rhythmically integrated. Many additional transitions could be perceived as rhythmic, but did not show sufficient isochronony in the acoustic analysis.[3] The majority of rhythmic turn transitions were either monosyllabic turns (such as "yeah" or "no") or turns in which the first syllable was rhythmically integrated, and the speaker then continued with a more syllable-timed rhythm.

This suggests that in spite of considerable differences in speech rhythm, at the point where it matters most, SE speakers often accomplish interactionally what their "native-speaking" counterparts accomplish, i.e., smooth transitions from one speaker to the next. In order to do so it is not necessary for them to speak with

stress-timed rhythm, but only to perceive stress timing in their British English speaking co-participants, and to orient to it wherever this becomes relevant in interaction, i.e., at the point of turn transition. Thus, SE speakers show interactional competence without adhering to BE pronunciation rules. From a conversational perspective it is not important how stress-timed or syllable-timed learners' speech is, but how successfully they employ rhythm for the accomplishment of conversational actions, such as turn-taking.

Concluding observations

The discursive perspective on pronunciation has gained much ground in recent years and will continue to do so with the increase in research on the role of phonetics and prosody for interaction. Insights from discourse and conversation analysis that reveal talk to be a collaborative achievement, rather than an individualistic activity, have filtered into much of communicative language teaching practice, and concepts that used to be considered the exclusive responsibility of individual speakers are now addressed as interactional issues. This applies, for example, to the concept of fluency, which McCarthy (2009) suggests to consider

> ... as an interactive achievement, perhaps more adequately captured by the metaphor of confluence. Achieving confluence, successfully interacting in talk that flows and being perceived as both able to create within one's own utterances and across utterances the satisfactory perception of flow for all participants is an art, the evidence of which will not be found or fairly assessed in monologic contexts but in the robust evidence of dyadic and multi-party talk (2009: 23).

Similarly, pronunciation does not fall within the domain of the single speaker, as each utterance is designed for specific recipients in response to specific prior talk and in order to accomplish a social action fitted to its specific context. Furthermore, as Lindemann (2006, 2011) has shown, intelligibility is as much the responsibility of the listener as it is that of the speaker. Therefore the teaching and learning of pronunciation requires an understanding of its nature as fundamentally entwined with the collaborative activity that is talk-in-interaction.

NOTES

1 Excerpts labelled SBC are taken from the Santa Barbara Corpus of Spoken American English (SBCSAE), a collection of naturally occurring spoken language data (Du Bois et al. 2000, 2003; Du Bois and Englebretson 2004, 2005). Recordings have been obtained from www.talkbank.org (MacWhinney 2007).

2 Syllables may start with a vowel or with one, two, or three consonants, which means the time it takes to articulate syllable onsets varies greatly.

3 Auer et al. (1999) provide no statistical information on how often rhythmic integration occurs in their native speaker data and they do not use the same rigorous method of measurement applied here. Instead, Auer et al. used physical methods such as tapping along to the speech in order to establish whether a transition was rhythmically integrated or not. Further, they suggest that nonintegration occurs when next speakers have problems with a previous turn. There can therefore be no expectation on L2 speakers to produce rhythmic integration at every turn transition.

REFERENCES

Abercrombie, D. 1967. *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.

Adams, C. 1979. *English Speech Rhythm and the Foreign Learner*, The Hague: Mouton.

Anderson-Hsieh, J., Johnson, R., and Koehler, K. 1992. The relationship between native speaker judgements of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning* 42(4): 529–555.

Anderson-Hsieh, J. and Venkatagiri, H. 1994. Syllable duration and pausing in the speech of Chinese ESL speakers. *TESOL Quarterly* 28(4): 807–812.

Auer, P., Couper-Kuhlen, E., and Müller, F. 1999. *Language in Time: The Rhythm and Tempo of Spoken Interaction*, New York: Oxford University Press.

Benton, M., Dockendorf, L., Jin, W., Liu, Y., and Edmondson, J.A. 2007. The continuum of speech rhythm: computational testing of speech rhythm of large corpora from natural Chinese and English speech. In: *Proceedings of the 16th International Congress of Phonetic Sciences.* http://www.icphs2007.de/conference/Papers/1591/1591.pdf.

Bond, Z.S. and Fokes, J. 1985. Non-native patterns of English syllable timing. *Journal of Phonetics* 13: 407–420.

Brown, A. 1988. The staccato effect in the pronunciation of English in Malaysia and Singapore. In: *New Englishes: The Case of Singapore*, J. Foley (ed.), 115–147, Singapore: Singapore University Press.

Chen, Y., Robb, M.P., Gilbert, H.R., and Lerman, J.W. 2001. A study of sentence stress production in Mandarin speakers of American English. *Journal of the Acoustical Society of America* 109(4): 1681–1690.

Couper-Kuhlen, E. 1996. The prosody of repetition: on quoting and mimicry. In: *Prosody in Conversation*, E. Couper-Kuhlen and M. Selting (eds.), 366–405, Cambridge: Cambridge University Press.

Couper-Kuhlen, E. 2004. Prosody and sequence organization: the case of new beginnings. In: *Sound Patterns in Interaction. Cross-linguistic Studies from Conversation*, E. Couper-Kuhlen and C.E. Ford (eds.), 335–376, Amsterdam: Benjamins.

Dauer, R.M. (1983). Stress-timing and syllable-timing reanalysed. *Journal of Phonetics* 11: 51–62.

Deterding, D. 2001. The measurement of rhythm: a comparison of Singapore and British English. *Journal of Phonetics* 29: 217–230.

Du Bois, J.W. and Englebretson, R. 2004. *Santa Barbara Corpus of Spoken American English, Part 3*, Philadelphia: Linguistic Data Consortium.

Du Bois, J.W. and Englebretson, R. 2005. *Santa Barbara Corpus of Spoken American English, Part 4*, Philadelphia: Linguistic Data Consortium.

Du Bois, J.W., Chafe, W,L., Meyer, C., and Thompson, S.A. 2000. *Santa Barbara Corpus of Spoken American English, Part 1*,

Philadelphia: Linguistic Data Consortium.
Du Bois, J.W., Chafe, W.L., Meyer, C., Thompson, S.A., and Martey, N. 2003. *Santa Barbara Corpus of Spoken American English, Part 2*, Philadelphia: Linguistic Data Consortium.
Dziubalsak-Kołaczyk, K. and Przedlacka, J. (eds.). (2005. *English Pronunciation Models: A Changing Scene*, Bern: Peter Lang.
Ford, C. E. and Thompson, S.A. 1996. Interactional units in conversation: syntactic, intonational, and pragmatic resources for the projection of turn completion. In: *Interaction and Grammar*, E. Ochs, E.A. Schegloff, and S.A. Thompson (eds.), 135–184, Cambridge: Cambridge University Press.
French, P. and Local, J. 1986. Prosodic features and the management of interruptions. In: *Intonation and Discourse*, C. Johns-Lewis (ed.), 157–180, London: Croom Helm.
Jenkins, J. 2000. *The Phonology of English as an International Language*, Oxford: Oxford University Press.
Klewitz, G. and Couper-Kuhlen, E. 1999. Quote-unquote. The role of prosody in the contextualization of reported speech sequences. *Pragmatics* 9(4): 459–485.
Levis, J.M. (ed.) 2005. Reconceptualizing Pronunciation in TESOL: Intelligibility, Identity, and World Englishes, Special issue. *TESOL Quarterly* 39(3).
Lindemann, S. 2006. What the other half gives: the interlocutor's role in non-native speaker performance. In: *Spoken English, TESOL and Applied Linguistics: Challenges for Theory and Practice*, R. Hughes (ed.), 23–49, Basingstoke: Palgrave Macmillan.
Lindemann, S. 2011. Who's "unintelligible"? The perceiver's role. *Issues in Applied Linguistics* 18(2): 223–232.
Local, J. 1992. Continuing and restarting. In: *The Contextualization of Language*, P. Auer and A. di Luzio (eds.), 273–296, Amsterdam: John Benjamins.
Local, J., Kelly, J., and Wells, B. 1986. Towards a phonology of conversation: turn-taking in Tyneside English. *Journal of Linguistics* 22(2): 411–437.
Local, J., Wells, B., and Sebba, M. 1985. Phonology for conversation: phonetic aspects of turn delimitation in London Jamaican. *Journal of Pragmatics* 9: 309–330.
Low, E.L. 2006. A cross-varietal comparison of deaccenting and given information: implications for international intelligibility and pronunciation teaching. *TESOL Quarterly* 40(4), 739–761.
Low, E.L., Grabe, E., and Nolan, F. 2000. Quantitative characterizations of speech rhythm: syllable-timing in Singapore English. *Language and Speech* 43(4): 377–401.
MacWhinney, B. 2007. The TalkBank project. In: *Creating and Digitizing Language Corpora: Synchronic Databases*, J.C. Beal, K.P. Corrigan, and H.L. Moisl (eds.), vol.1, 163–180, Basingstoke: Palgrave-Macmillan.
McCarthy, M. 2009. Rethinking spoken fluency. *Estudios de Lingüística Inglesa Aplicada* 9: 11–29. http://institucional.us.es/revistas/elia/9/3.%20McCarthy.pdf.
Miller, M. 1984. On the perception of rhythm. *Journal of Phonetics* 12, 75–83.
Pike, K.L. 1945. *The Intonation of American English*, Ann Arbor, MI: University of Michigan Publications.
Reber, E. 2012. *Affectivity in Interaction. Sound Objects in English*, Amsterdam: John Benjamins.
Sacks, H., Schegloff, E.A., and Jefferson, G. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language* 50: 696–735.
Schegloff, E.A. 2007. *Sequence Organization in Talk-in-Interaction: A Primer in Conversation Analysis*, Cambridge: Cambridge University Press.
Selting, M. 1996. Prosody as an activity-type distinctive cue in conversation: the case of so-called "astonished" questions in repair initiation. In: *Prosody in Conversation*, E. Couper-Kuhlen and M. Selting (eds.), 231–270, Cambridge: Cambridge University Press.

Selting, M., Auer, P., Barden, B., Bergmann, J.R., Couper-Kuhlen, E., Günthner, S., Meier, C., Quasthoff, U., Schoblinski, P., and Uhmann, S. 1998. Gesprächsanalytisches Transkriptionssystem (GAT). *Linguistische Berichte* 173: 91–122.

Szczepek Reed, B. 2004. Turn-final intonation in English. In: *Sound Patterns in Interaction. Cross-linguistic Studies from Conversation*, E. Couper-Kuhlen and C.E. Ford (eds.), 97–118, Amsterdam: Benjamins.

Szczepek Reed, B. 2006. *Prosodic Orientation in English Conversation*, Basingstoke: Palgrave.

Szczepek Reed, B. 2009. Prosodic orientation: a practice for sequence organization in broadcast telephone openings. *Journal of Pragmatics* 41(6): 1223–1247.

Szczepek Reed, B. 2010. Speech rhythm across turn transitions in cross-cultural talk-in-interaction. *Journal of Pragmatics* 42(4), 1037–1059.

Szczepek Reed, B. 2012a. Beyond the particular: prosody and the coordination of actions. *Language and Speech* 55(1): 12–33.

Szczepek Reed, B. 2012b. A conversation analytic perspective on teaching English pronunciation: the case of speech rhythm. *International Journal of Applied Linguistics* 22(1): 67–87.

Taylor, D.S. 1981. Non-native speakers and the rhythm of English. *International Review of Applied Linguistics* 19(3): 219–226.

Wells, B. and Peppe, S. 1996. Ending up in Ulster: prosody and turn-taking in English dialects. In: *Prosody in Conversation*, E. Couper-Kuhlen and M. Selting (eds.), 101–130, Cambridge: Cambridge University Press.

Appendix

Transcription Conventions (adapted from Selting et al. 1998)

Pauses and lengthening

(2.85)	measured pause
:::	lengthening

Accents

ACcent	primary pitch accent
Accent	secondary pitch accent

Phrase-final pitch movements

?	rise-to-high
,	rise-to-mid
-	level
;	fall-to-mid
.	fall-to-low

Pitch step-up/step down

↑	pitch step-up
↓	pitch step-down

Changes in pitch register and volume

<<l> >	low pitch register
<<h> >	high pitch register
<<f> >	forte
<<p> >	piano

Breathing

.h, .hh, .hhh	in-breath
h, hh, hhh	out-breath

Other conventions

[	overlapping talk
[	

12 Fluency

RON I. THOMSON

Introduction

As a lay term, *fluency* is often used to denote general second language (L2) proficiency. In this context, the term typically implies that an L2 user has advanced facility with the grammar, vocabulary, and perhaps even the pronunciation of a second language (Segalowitz 2010). The term fluency might also be used to indicate that a person can comprehend the L2 with ease or that the person has advanced skills in L2 reading and writing. Notably, this lay use of the term necessarily excludes its application to learners who are beginners and even to those with an intermediate knowledge of an L2.

In contrast, applied linguists and language teachers typically use the term fluency to refer to the fluidity or ease with which the second language is spoken (Derwing et al. 2004; Freed 2000; Isaacs and Thomson 2013; Koponen and Riggenbach 2000). Consequently, some lower proficiency L2 learners may be described as fluent, despite the fact that they have only rudimentary grammatical ability, limited vocabulary knowledge, and poor pronunciation. In this context, describing lower proficiency learners as fluent is understood to mean that the language knowledge they do have is easily accessed and that their oral language is produced without undue hesitation (Segalowitz 2010). At the extreme, pidgin languages provide an example of second language varieties that develop into highly fluent systems of communication, despite comprising substantially reduced morphological, grammatical, lexical, and phonological forms (Holm 2000).

Anyone in the field of second language instruction will have encountered learners who produce fluent but structurally simplified L2 speech at the same proficiency level as learners who, despite having similar declarative knowledge, are disfluent. These individual differences in oral performance across speakers with similar knowledge of a language are widely assumed to emerge from a trade-off between accuracy and fluency, whereby L2 learners' attention to form can adversely affect their fluency (Skehan 1998, 2009; VanPatten 1990). Competition

The Handbook of English Pronunciation, First Edition. Edited by Marnie Reed and John M. Levis.

between accuracy and fluency arises when learners access declarative knowledge during online processing, since doing so requires reliance on short-term memory, which is assumed to have a limited capacity (Baddeley 2007). When a speaker's conscious attention is directed toward one part of the speech production process, such as pronunciation, less attention is available for other processes, such as lexical access, grammatical encoding, etc. This can lead to a breakdown in the speech production system, manifesting itself as a disfluent utterance.

In this chapter, I endeavor to disentangle fluency as a cognitive skill from other constructs of speech production commonly found in the pronunciation literature, and review the few studies that hint at a relationship between fluency and three of these constructs – accentedness, intelligibility, and comprehensibility. I then describe some theoretical models that help illuminate the role of pronunciation in the development of oral fluency. Finally, I discuss some implications for pronunciation instruction.

Defining fluency

In the pronunciation literature, fluency is often considered in combination with other measures of spoken language – especially comprehensibility and accentedness (e.g., Derwing and Munro 1997; Derwing, Munro, and Thomson 2008; Derwing, Thomson, and Munro 2006; Isaacs and Thomson 2013). However, although they are often discussed together, and despite superficial similarities in how they are measured, fluency is quite different from these other constructs.

In accentedness, intelligibility, and comprehensibility research, the interaction between L2 speakers' production and listeners' perception forms the locus of attention. For example, accentedness is operationalized using impressionistic judgments of how far L2 speakers' pronunciation diverges from a native speaker target; intelligibility is operationalized in terms of how accurately listeners are able to identify spoken language relative to an L2 speaker's intended utterance; and comprehensibility is operationalized as how easy speech is for a listener to understand, referring to how much effort is involved (see Munro and Derwing 1995a; Munro, Derwing, and Morton 2006).

In contrast, investigations of oral fluency typically focus on the state of learners' L2 speech production systems. Thus, although measures of oral fluency often involve listener judgments, those judgments are typically understood to reflect the underlying cognitive processes involved in planning and producing spoken language, and the degree to which those processes are automatic or controlled. Listeners' perceptions of L2 fluency in terms of its impact on communication are normally of little interest in this line of research.

Although listener judgments are taken to be indicators of fluency, other features in learner speech undoubtedly influence these judgments (e.g., word choice, grammar, pronunciation). Recognizing this problem, Derwing et al. (2004) blend definitions from Schmidt (1992) and Guillot (1999) to describe fluency as comprising "an automatic procedural skill on the part of the speaker and a perceptual phenomenon in the listener" (2004: 656).

Given that listener judgments of cognitive fluency can be influenced by other factors, some researchers have attempted to operationalize fluency using more objectively quantifiable correlates of fluency, which are extracted from the speech signal itself (see Derwing et al. 2004, 2009; Kormos 2006; Kormos and Dénes 2004; Towell, Hawkins, and Bazergui 1996). Among others, these measures often include:

- Speech rate: the average number of syllables spoken per second or minute.
- Phonation time ratio: the percentage of time devoted to speaking relative to the total time taken to produce an utterance.
- Pruned syllables: the average number of syllables spoken per second or minute after any disfluencies have been removed (e.g., syllables attributed to self-repetitions, self-corrections, etc., are not counted).
- Articulation rate: the average number of fluent syllables per second or minute between pauses of a predetermined length (e.g., 400 ms). Like pruned syllables, this measure excludes disfluencies from the total syllable count. Unlike pruned syllables, when calculating the duration of the utterance, any time elapsed during the production of disfluencies (e.g., self-repetitions) and pauses is excluded from the total time.
- Mean length of run: the average number of words or syllables produced between pauses of a length specified by the researcher(s).
- Silent pause ratio: the number and/or time attributed to silent pauses of a particular length per second or minute. The minimum duration for what constitutes a silent pause varies across studies.
- Filled pause ratio: the number and/or duration of pausing attributed to filled pauses (e.g., "um" and "uh") per second or minute.

In an attempt to overcome some of the limitations associated with using either listener judgments or temporal measurements on their own, Derwing et al. (2009) employed both methods of assessment in a longitudinal study aimed at examining the link between L1 and L2 fluency. Interestingly, while some of the listener judgments closely paralleled the related temporal measures, on other occasions this was not the case. Overall, temporal measures were found to be more sensitive in detecting a relationship between L1 and L2 fluency than were listener judgments of the same speech samples. The authors attributed this difference to the fact that the judges who assessed the L1 samples were not the same judges who assessed the L2 samples. However, one might just as reasonably conclude that temporal measures are simply more accurate than listener judgments, which are influenced by other unrelated factors.

As Segalowitz (2010) points out, notwithstanding the additional insight temporal measures offer relative to listener judgments, operationalizing fluency in these terms is not entirely satisfactory. For example, researchers make subjective judgments in concluding that every self-repetition or self-correction in a given utterance is a sign of cognitive disfluency (Hieke 1981; MacGregor, Corley, and Donaldson 2009). In fact, speakers sometimes use self-repetitions and corrections as a discourse strategy, aimed at clarifying or emphasizing given information for the listener's benefit.

Similarly, deciding what length of silent pause constitutes a disfluency, and in what context, is also subjective (see Davies 2003). Like self-repetition, pausing can be used as a discourse strategy, especially at clause and sentence boundaries. As a result, silent pauses do not provide fail-safe evidence that there has been a breakdown in cognitive fluency. Conversely, lexical filled pauses (e.g., "like", "I think", etc.), which often are a sign of disfluency, are never considered when the filled pause ratio is calculated. This omission is problematic, since like their non-lexical counterparts, lexical filled pauses are sometimes produced in order to buy time for planning and producing utterances that follow.

Defining fluency in terms of temporal phenomena can also be limiting because it reduces the construct to speed of production, which is an oversimplification of the complex cognitive underpinnings of fluency. Taking a cognitive perspective, Schmidt (1992) and Segalowitz (2010) argue that it is the efficiency and automaticity of processing, rather than speech rate, that marks fluency. This means that while, on average, automatic processing is likely to be faster than conscious processing, speech rate cannot serve as a reliable indicator of fluency. Instead, fluency should be viewed as a set of interrelated and overlapping cognitive processes, organized in such a way that they impose the smallest possible demands on working memory, which has limited capacity.

Although a more transdisciplinary approach using experimental techniques borrowed from psychology and neuroscience would unquestionably provide more precision in measuring cognitive fluency (see Segalowitz 2010 for a detailed overview), a substantial body of research using less sophisticated approaches has still provided important insights into the development of L2 fluency. It is to this earlier research that we next turn.

Relationships between fluency, accentedness, intelligibility, and comprehensibility

While interrelations between accentedness, intelligibility. and comprehensibility are well attested, their relationship to fluency is less so. To date, no studies appear to have systematically investigated the relationship between oral fluency and these other commonly researched measures of L2 speech production. This is somewhat surprising, since learners' conscious attention to pronunciation can affect their fluency, which in turn might impact listener perceptions of accent, intelligibility and comprehensibility. Despite the paucity of deliberate research in this area, several studies have examined fluency and these other measures of oral production in tandem. By re-examining several such studies, preliminary evidence emerges that fluency is partially related to these other facets of L2 pronunciation.

Accentedness, intelligibility, and comprehensibility

Before exploring the relationship between fluency and other dimensions of L2 speech, it is helpful to have a basic understanding of the research contrasting accentedness with intelligibility and comprehensibility.

Munro and Derwing (1995a) were the first to empirically demonstrate that an L2 speaker could have a very strong foreign accent, but still be highly intelligible and comprehensible. In a later study, Derwing and Munro (1997) confirmed that this quasi-independence of accentedness, intelligibility, and comprehensibility extends across proficiency levels and across learners from varied L1 backgrounds. Furthermore, while the native speaker raters who participated in the later study reportedly recognized that non-target-like segmental features contributed to their perception of a foreign accent, the raters did not report the same influence on their comprehensibility ratings. Derwing and Munro interpreted this to mean that while segmental errors are a major contributor to the perception of a foreign accent, they do not necessarily lead to processing difficulties for listeners. To determine whether impressionistic ratings of accentedness are aligned with more objective measures of comprehensibility, Munro and Derwing (1995b) examined the relationship between perceived accentedness and the amount of time it took for listeners to process the accented utterances. This study confirmed that a strong foreign accent does not always lead to an increase in processing time.

Derwing, Munro, and Wiebe (1998) extended these findings to classroom instruction by investigating whether 12 weeks of pronunciation instruction focused on segmentals (i.e., vowels and consonants) versus instruction focused on suprasegmentals (e.g., word stress, intonation, rhythm) would have a greater impact on listener ratings of L2 speech. They found that suprasegmental training led to significant improvement in comprehensibility ratings of speech during an extemporaneous speaking task, while segmental training did not. In read speech, both groups showed improvement. Since the ultimate goal of pronunciation instruction is to improve the intelligibility and comprehensibility of spontaneous production, Derwing, Munro, and Wiebe (1998) argue that focusing on suprasegmental features provides the greatest benefit to learners.

Fluency and accentedness

Indirect evidence for a link between fluency and accentedness can be found by considering other measures to which these constructs are both related. For example, Derwing et al. (2004) found that speech rate and goodness of prosody (referring to suprasegmental features) are correlated with fluency, while other studies report that speech rate and prosody are correlated with accentedness (e.g., Derwing and Munro 1997; Kang 2010). In another study, Trofimovich and Isaacs (2012) report a moderate negative correlation between mean length of run, a measure of fluency, and number of segmental errors, a measure typically associated with accentedness.

Derwing and Rossiter (2003) investigated the fluency, phonological accuracy, and complexity of L2 speech before and after pronunciation training. They found that the phonological accuracy of learners receiving segmental instruction significantly improved, even on an extemporaneous speaking task, while these learners experienced no improvement in terms of fluency and complexity. This was in marked contrast to a group trained on suprasegmentals, for whom fluency and complexity significantly improved, while their phonological accuracy did not. The

researchers concluded that the segmental group had consciously attended to phonological form during their speech production and that this had consumed cognitive resources that could otherwise have been used for more fluent and complex speech. In fact, while there may have been no improvement in the segmental groups' fluency over time, neither was there a significant decline. If conscious attention to phonological form impacted other processes involved in speech production, as the authors argue, we should expect that the speakers' fluency and complexity scores would decline as a result. The fact that this did not happen hints at the possibility that real gains in fluency and complexity were masked by the deleterious effect of the learners attending to phonological form. This interpretation suggests that as the segmental groups' newly learned phonological knowledge becomes automatized, their fluency scores might also improve.

Derwing, Thomson, and Munro (2006) examined Mandarin and Slavic speakers' development of L2 English fluency and accentedness over a period of eight

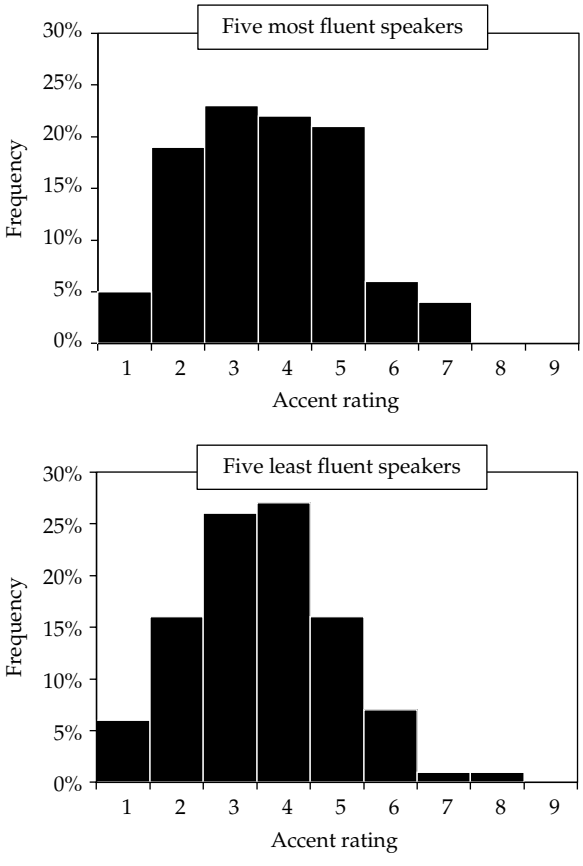


Figure 12.1 Distribution of accent ratings for the five most fluent and five least fluent speakers (1 = strong accent; 9 = no accent).

months. Regrettably, the researchers do not report any correlational analysis between the fluency and accentedness ratings they obtained. Nevertheless, we could surmise that if fluency and accentedness are related, improvement in one should be accompanied by improvement in the other. This prediction is only partially borne out in this study. While both fluency and accent ratings improved for the Slavic group, only accent ratings improved for the Mandarin group. However, since the absolute increase in the Mandarin group's accent ratings was quite small, this could be taken to indicate that greater improvement in pronunciation is necessary before a relationship with fluency becomes detectable.

A colleague and I recently examined the impact of rater expertise and rating scale length on listener judgments of the fluency, accentedness, and comprehensibility of 38 L2 English learners (Isaacs and Thomson 2013). For the purpose of this chapter, I have revisited that data to explicitly examine the relationship between the fluency and accentedness ratings for the 20 raters who used a 9-point scale. A Pearson's r coefficient reveals a moderate correlation between these two constructs ($r = 0.605$, $p < 0.001$). To probe further, I examined the distribution of accentedness ratings for the five most fluent and five least fluent speakers respectively (see Figure 12.1).

Although skewed toward the low end of the scale (i.e., away from a native-like accent), the distributions are otherwise relatively normal. This indicates that it is sometimes possible for a speaker to be perceived as more fluent, but highly accented, or conversely as less fluent despite having a more native like accent.

Fluency and intelligibility

Previous studies appear to provide only limited evidence for a relationship between fluency and intelligibility. For example, Munro and Derwing (1995a) investigated a number of variables with the intent of establishing an error gravity hierarchy for intelligible L2 speech. They concluded that prosodic errors have a more negative impact on intelligibility than most other error types. In a later study, Derwing and Munro (1997) report that goodness of prosody, as well as speech rate, were significantly correlated with the intelligibility scores of a small subset (8%) of listeners in their study. Since goodness of prosody and speech rate are also known to be moderately correlated with fluency ratings (Derwing et al. 2004), these studies provide an indication that fluency and intelligibility may be weakly related. Relationships between prosody and fluency might suggest that intelligibility can be improved through fluency instruction or that instruction in prosody could impact fluency.

Fluency and comprehensibility

Indirect evidence for a relationship between fluency and comprehensibility can be found across several studies. For example, Anderson-Hsieh, Johnson, and Koehler (1992) report that accurate prosody is linked to more comprehensible speech. Similarily, Derwing and Munro (1997) report that goodness of prosody and speech rate were correlated with the comprehensibility ratings of a third of the raters in

their study. In addition, 15% of their raters indicated that they were consciously aware that several features normally associated with fluency (e.g., pausing, speech rate, etc.) had affected their comprehensibility judgments.

The pronunciation training studies described previously (i.e., Derwing, Munro and Wiebe 1998; Derwing and Rossiter 2003) also point to an indirect relationship between fluency and comprehensibility. While the group trained on suprasegmentals experienced significant improvement in both dimensions on the extemporaneous speech task, the group trained on segmentals did not. This suggests that fluency and comprehensibility are both closely aligned with suprasegmental features of pronunciation.

Derwing, Munro, and Thomson (2008) compared the development of Mandarin and Slavic learners' English fluency and comprehensibility over a two-year period. Although the relationship between these two constructs was not a focus of their study, they report moderate to strong correlations between fluency and comprehensibility ratings at three separate data collection points (Pearson's *r* coefficients were 0.872, 0.791, and 0.791 respectively).

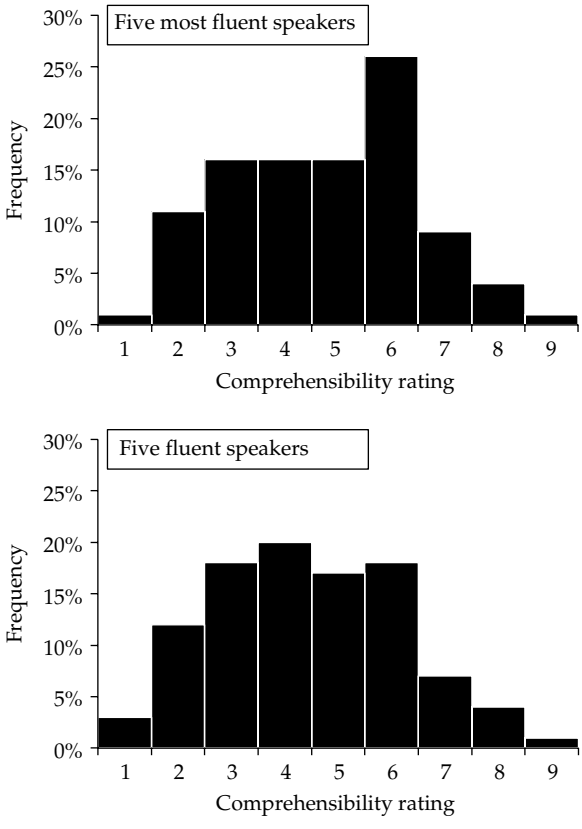


Figure 12.2 Distribution of comprehensibility rating for the five most fluent and five least fluent speakers (1 = extremely difficult to understand; 9 = very easy to understand).

Revisiting data from Isaacs and Thomson's (2013) study reveals a strong correlation between fluency and comprehensibility ratings ($r = 0.826$, $p < 0.001$). However, comprehensibility ratings for the five most fluent and five least fluent speakers (see Figure 12.2) are quite normally distributed. This indicates that it is possible for some L2 speakers to be perceived as more fluent, but less comprehensible, or conversely as less fluent, but highly comprehensible.

Summary

Taken together, the findings summarized in this section provide evidence that fluency is most related to comprehensibility, somewhat related to accentedness, and apparently least related to intelligibility. In the latter regard, the evidence is admittedly quite limited. These patterns have important implications for instruction, since they suggest that improvement in fluency may lead to improvement in comprehensibility and accentedness. Conversely, improvement in prosody and segmental accuracy might also lead to improvement in fluency.

Relevant speech production models

Several theoretical models of speech production help illuminate the complex interactions between fluency and pronunciation. Specifically, these models help point to possible underlying cognitive mechanisms and processes that might explain how improvement in either fluency or pronunciation can promote improvement in the other.

Adaptive Control of Thought model

Although some of its general assumptions about language are open to debate (Towell, Hawkins, and Bazergui 1996), Anderson's (1983, 1993) Adaptive Control of Thought (ACT) model provides a useful framework for explaining differences between controlled and automatic processing in speech production. These differences are said to account for the trade-off between fluency and accuracy (e.g., Derwing and Rossiter 2003; Skehan 2009). ACT has earlier been used to explain the development of L2 fluency (e.g., Segalowitz 2010; Towell, Hawkins, and Bazergui 1996), but has not previously been extended to a specific discussion of fluency's relationship to pronunciation.

ACT divides memory into three subtypes: declarative memory, production memory, and working memory. Declarative memory comprises long-term knowledge that must be consciously retrieved prior to use, while production memory contains long-term knowledge that is automatically retrieved without conscious attention. In contrast, working memory is a short-term memory store, which briefly holds small amounts of information retrieved from one of the two long-term memory stores during online speech production. Working memory is also used to temporarily hold new information encountered in the outside world, before it can be added to long-term memory. According to ACT, the movement of

information from declarative memory to production memory is also mediated by working memory.

ACT takes the consensus view that retrieval of declarative knowledge is inefficient relative to retrieval of procedural knowledge. In part, this stems from an understanding that the working memory component has limited capacity – see, for example, the widely cited Miller's law (Miller 1956), which posits a human capacity to hold seven plus or minus two units of information in short-term memory. Access to and manipulation of declarative knowledge during speech production places a high demand on working memory, both in terms of its capacity and the amount of attention it requires. In contrast, procedural knowledge consumes far less memory and attention since its units, stored in production memory, partially comprise preprocessed chunks of information. For example, at the lexical level, procedural knowledge may include automatized collocational connections between smaller units and storage of prefabricated lexical chunks and phrases. ACT describes connections between smaller units within production memory as IF/THEN pairs. These pairs are accessed nearly simultaneously because activation of the first half of the pair automatically activates the other half. Since each resulting unit of speech moving from production memory to working memory is larger than those from declarative memory, the entire system is more efficient, which promotes greater fluency.

Applying ACT to pronunciation, phonological forms for associated units (e.g., words or phrases) can be retrieved from either declarative or production memory. For normal adult speakers, L1 pronunciation constitutes procedural knowledge, and is therefore automatic. In L2 speech, a learner can still rely on automatized L1 phonological processes, which might allow the learner to be more fluent, albeit highly accented. However, the development of more accurate L2 phonology begins within declarative and working memory stores. Thus, the only way to offset the impact of attention to phonological form on fluency would be to make other parts of the process automatic, freeing up more processing capacity to attend to pronunciation.

In order for a pronunciation correction to take place before an L2 utterance is spoken, speakers must consciously access declarative knowledge. Because this strategy is inefficient, it can result in disfluent speech. Disfluencies arising from reliance on declarative knowledge might simply be manifested as temporal hesitation. This would make identifying the source of the disfluency difficult, leading to a weak correlation with pronunciation. In contrast, adjustment to pronunciation during automatic speech production can only take place after a learner has already heard his or her own utterance and perceives a mismatch between that utterance and explicit knowledge. In such cases, since the utterance has already been spoken, the only option for a repair is to make a self-correction. This explicitly implicates pronunciation as the source of the disfluency. Even when new L2 pronunciation patterns move into production memory, the speed with which they are accessed might remain slow until connections between these forms and other knowledge (e.g., the words in which they occur) are strengthened. Despite being slower, because they are automatic, they do not represent a

Table 12.1 Possible learner outcomes for L2 fluency and pronunciation applying Anderson's (1983) Adaptive Control of Thought model.

	Uses L1 phonology	*Uses interlanguage phonology*	*Uses L2 phonology*
Production memory	Fluent – Accented • High risk to intelligibility and comprehensibility	Fluent – Somewhat accented • Potential risk to intelligibility and comprehensibility	Fluent – Unaccented • No risk to intelligibility and comprehensibility
Declarative memory	Disfluent – Accented • High risk to intelligibility and comprehensibility	Disfluent – Somewhat accented • Potential risk to intelligibility and comprehensibility	Disfluent – Unaccented • No risk to intelligibility; some to comprehensibility

disfluency (see Segalowitz 2010). Nevertheless, speed of access does influence traditional fluency measures.

ACT's assumptions about long- and short-term memory allow us to make specific predictions regarding the relationship between L2 fluency and pronunciation. Table 12.1 highlights six possibilities, depending on the state of the L2 learner's phonological system. It predicts, for example, that learners can be heavily accented, but still fluent, if they use the already established L1 phonological system that is part of their production memory. In contrast, too much attention to accurate pronunciation of the L2 might lead to disfluent speech. It might also be possible for learners to consciously match L1 sounds to L2 vocabulary, in which case they would be both accented and disfluent.

Levelt's model of Speech Production

Levelt's (1989, 1999) "blueprint" of speech production provides another useful framework for discussing the relationship between L2 fluency and pronunciation. This widely cited model has been adapted to describe L2 speech production by De Bot (1992), further elaborated by others (Kormos 2006; Segalowitz 2010; Skehan 2009). The purpose here is not to describe Levelt's model in detail, but to highlight how it can be used to explain relations between fluency and pronunciation. It complements Anderson's (1983) ACT model by describing the processes involved in speech production, rather than focusing on the role of memory.

In brief, Levelt's (1999) model describes the speaking process as primarily linear, although it does allow a few iterative steps. The first step in the model is the conceptual preparation of a pre-verbal message. This is followed by processes

involved in selecting what grammatical and rhetorical forms should be used to convey that message. Next are procedures that apply any rules necessary for the speaker to arrive at the correct spoken form of the utterance. In terms of pronunciation, these rules specify how a word or utterance is to be articulated. Like Anderson's (1983) description of production memory, many steps in Levelt's model overlap, with one step in the process simultaneously activating the following step. Levelt's model is somewhat predictive in that for reasons of efficiency particular steps of the process anticipate what will come later. The process in Levelt's model can be interrupted or slowed at numerous junctions, resulting in disfluency. In Anderson's (1993) terms, these junctions might represent points in the process where the speaker moves between productive memory, with its automatic processing, and declarative memory, with its controlled processing.

While many disfluency-invoking breakdowns in Levelt's Speech Production model are unrelated to pronunciation (e.g., searching for a word or grammatical form), in other cases pronunciation difficulties might very well be the source of temporal or perceived disfluencies (see Segalowitz 2010 for a detailed discussion of specific points in Levelt's model where disfluency may be expected to arise). The first place in the system where pronunciation may cause a breakdown is when phonological encoding is applied to a planned utterance. At this point, a speaker accesses the mental lexicon to assign segmental features, syllabification of words, and prosody at the phrasal level. In Levelt's model, the mental lexicon comprises implicit lexical and morphological knowledge, which is automatic. In contrast, except for very advanced speakers, L2 words largely comprise declarative knowledge, because these words are not yet fully established in the mental lexicon, either in terms of access or in terms of their connections to other words. Like Anderson's (1993) ACT, in Levelt's model accessing declarative knowledge requires conscious attention. One obvious way to compensate for this added demand on the speech production system is to link L2 vocabulary with L1 phonology, as many L2 speakers do. This will result in the fluent–accented speaker illustrated in Table 12.1.

After phonological encoding, the second location in Levelt's blueprint where pronunciation may impact fluency is during phonetic encoding. During phonetic encoding, the input from the phonological encoding process must activate related articulatory gestures. For normal adult L1 speech, these gestures are largely automatic. Thus, an L2 speaker who uses L1 gestures will fall into the fluent–accented category. Conversely, while L2 gestures will rarely be native-like, some interlanguage gestures may become automatized and therefore lead to the fluent–somewhat accented category of Table 12.1. If, however, a learner accesses declarative knowledge for L2 gestures, whether accurate or not, this might lead to the disfluent–somewhat accented category.

The third place for a possible influence of pronunciation on fluency is when the speaker leaves the phonetic encoding stage and enters the articulation stage. Here, all the planning involved in the previous stages results in overt speech. The extent to which L1, L2, or interlanguage articulatory procedures are automatized will impact fluency in much the same way as in the previous stage. The automatic

nature of the process can further break down if the learner, while monitoring his or her overt speech, detects a need for repair in how he or she articulates an utterance. This can lead to a repair in the form of a self-correction.

Complexity Theory

While Anderson's (1983) ACT and Levelt's (1999) speech production model are helpful for elaborating the role of memory and automaticity in speech processing, Complexity Theory (CT) holds promise for describing how fluency and pronunciation interact and develop over time (see Larsen-Freeman 1997; Larsen-Freeman and Cameron 2008). Furthermore, since CT is transdisciplinary, it offers perspectives beyond those typically associated with structural linguistics, which, as Larsen-Freeman (2012) notes, has historically provided the foundation for our understanding of L2 speech production.

Several major tenets of CT have implications for descriptions of L2 fluency and pronunciation. For example, complex systems are said to be open and dynamic, implying that procedural knowledge *can* be changed. While this might seem to be at odds with the belief that L2 language skills become fossilized (e.g., Nakuma 1998; Selinker 1972), there is evidence that with the right sort of instructional intervention change in pronunciation is possible, even after learning is traditionally assumed to be asymptote (e.g., Derwing, Munro, and Wiebe 1997; Thomson 2011, 2012). CT's view that dynamic systems are open to change should not be interpreted as meaning that native-like pronunciation is attainable for all adult learners, but it clearly implies that given the right conditions at least some change is possible.

Another important principle from CT is that complex systems are emergent, and arise from multiple system components in interaction. Thus, while Anderson's (1993) and Levelt's (1989) models divide speech production into relatively discrete parts, CT assumes that changes in the functioning of one part of the system can impact other parts. This means that activities aimed at improving fluency could simultaneously improve pronunciation, and vice versa. Taking this further, improvement in both fluency and pronunciation might come from targeting a different part of the system altogether, for example, grammar or vocabulary.

Another important principle from CT is that there can be multiple routes to the same emergent system or outcome. This might explain why changes in fluency can impact changes in pronunciation and vice versa. More than ACT or Levelt's model, CT offers a framework for making sense of the sometimes chaotic evidence for a partial relationship between fluency, accentedness, intelligibility, and comprehensibility, and opens new directions for fluency and pronunciation research.

Implications

Although many classroom activities are purported to promote L2 fluency (see Rossiter et al. 2010), there is a dearth of research exploring their long-term impact. Related research has, however, revealed factors that affect oral fluency in the

short term. For example, many researchers have examined the impact of task type, planning, and rehearsal (e.g., Ellis 2009; Skehan and Foster 2008), while others have investigated the use of repetition (Gatbonton and Segalowitz 2005), time constraints (Nation 1989), and consciousness-raising (Boers et al. 2006). Unfortunately, the goal of such research is typically limited to validating theoretical models of speech production (e.g., De Bot 1992; Levelt 1999), with little attention to the influence of instructional practice on the development of fluency as a complex cognitive system. Thus, while a variety of factors can clearly be shown to impact fluency during a specific classroom task (state fluency), the extent to which these factors affect permanent changes in fluency (trait fluency) remain uncertain. Nevertheless, some general instructional principles can be inferred from what is known about the relationship between fluency and pronunciation, and through appeal to the theoretical models outlined in the previous section.

One general principle is that for pronunciation instruction to promote fluency, it should aim to stimulate transfer of declarative knowledge to production memory. This means that instructional activities should balance attention to phonological form with activities in which the same forms are represented in communicative contexts. For example, some activities could require learners to consciously attend to pronunciation, which would encourage more accurate but less fluent speech (e.g., Saito 2013). Other activities might use corrective feedback after an utterance has been spoken, since this will not interrupt the speech production process (e.g., Saito and Lyster 2012a, 2012b). Since opportunities to balance fluency and accuracy are rare outside of the classroom, where the demands of communication often prevent conscious attention to form (Lee et al. 1997; Schmidt 2001), this type of instruction is particularly important.

Skehan (2009) provides a useful summary of communicative classroom tasks that promote fluency and accuracy. Personal exchange tasks or other tasks with concrete and familiar topics promote both fluency and accuracy, since they place less of a burden on working memory. Similarly, tasks that provide clear structure also promote both fluency and accuracy. In contrast, tasks introducing new content, such as picture descriptions, or tasks requiring manipulation of information, are ill suited for fluency and pronunciation development, since they impose too many competing demands on working memory.

A second guiding principle for effective pronunciation instruction is that it should include activities that facilitate restructuring of the speech production process to make it more efficient. For example, in Levelt's (1999) model, a breakdown could first occur during phonological encoding of words. One strategy to reduce the potential for such breakdowns is to improve the overall speed of lexical access (see Skehan 2009). If words are accessed efficiently, more working memory capacity is available to devote to pronunciation. In fact, there is strong evidence that pronunciation accuracy is closely related to lexical frequency and familiarity (e.g., Munro and Derwing 2008; Thomson and Isaacs 2009; Walley and Flege 1999). Thus, vocabulary training and reinforcement should play a central role in fluency and pronunciation instruction.

Another place where restructuring could promote fluency is during the phonetic encoding stage of the speech production process. Connections between mental representations and articulatory gestures can be improved through explicit pronunciation instruction. Training ought to incorporate both perceptual and production practice, since it is widely assumed that in normal L2 development speech perception precedes speech production (Flege 2009). At the same time, in keeping with Complexity Theory, practice in producing sounds might offer another route to improving perception, a claim made by Lowie (2010). Reed and Michaud (2005) also appeal to this view in their argument that speaking helps listening, because learners' own speech becomes input in their developing L2 speech perception.

When deciding on the content of instruction, suprasegmental features should be given priority, since they are likely to impact fluency more than segmental features. When segmental features are taught, those that occur the most frequently in contrast with other sounds are likely to provide the greatest long-term benefit to fluency. The relative contribution of individual sounds to communication is known as their functional load (Brown 1991; Munro and Derwing 2006). Spending time on sounds with a low functional load can cause learners to unnecessarily divert attention toward features of pronunciation that do not have a major impact on the intelligibility of speech. Furthermore, when L1 sounds can be used in place of L2 sounds without a loss of intelligibility (e.g., a trilled /r/ instead of a standard English /r/), for reasons of efficiency there is merit in allowing learners to continue using the L1 sound rather than expecting them to acquire the more native-like form, if fluency is a goal.

As more accurate perceptual representations emerge, pronunciation instruction should provide learners with substantial practice in speech articulation at the level of segments, words, and phrases. This will promote fluency at the articulation stage of Levelt's (1999) model. As with the previous stage, it is advisable to allow learners to rely on L1 speech sounds, whenever doing so does not adversely affect their intelligibility. Unrealistically attempting to achieve accent-free production can lead learners to overmonitor their speech, causing at least temporary and unnecessary destabilization of an already efficient L1 system.

Conclusion

In this chapter I have related fluency to some common constructs from the pronunciation literature. This relationship can be further understood through reference to cognitive mechanisms that are known to impact fluency and pronunciation. Given the many questions that surround the validity of fluency measures used in existing L2 research, the precise nature of the relationship between fluency and pronunciation remains uncertain. Future research is needed that is more methodical in relating fluency to pronunciation. There is also a clear demand for longitudinal research in this area. This can lead to evidence-based pedagogical interventions, which will encourage both more fluent and more comprehensible L2 speech.

REFERENCES

Anderson J.R. 1983. *The Architecture of Cognition*, Cambridge, MA: Harvard University Press.

Anderson, J.R. 1993. *Rules of the Mind*, NJ: Lawrence Erlbaum Associates, Inc.

Anderson-Hsieh, J., Johnson, R., and Koehler, K. 1992. The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning* 42(4): 529–555.

Baddeley, A. 2007. *Working Memory, Thought, and Action*, Oxford: Oxford University Press.

Boers, F., Eyckmans, J., Kappel, J., Stengers, H., and Demecheleer, M. 2006. Formulaic sequences and perceived oral proficiency: putting a lexical approach to the test. *Language Teaching Research* 10(3): 245261. doi: 10.1191/1362168806lr195oa.

Brown, A. 1991. Functional load and the teaching of pronunciation. In: *Teaching English Pronunciation: A Book of Readings*, A. Brown (ed.), 221–224, London: Routledge.

Davies, A. 2003. *The Native Speaker: Myth and Reality*, 2nd edition, Clevedon: Multilingual Matters.

De Bot, K. 1992. A bilingual production model: Levelt's "Speaking" model adapted. *Applied Linguistics* 13: 1–24.

Derwing, T.M. and Munro, M.J. 1997. Accent, comprehensibility and intelligibility: evidence from four L1s. *Studies in Second Language Acquisition* 19: 1–16. doi: 10.1111/0023-8333.49.s1.8.

Derwing, T.M., Munro, M.J., and Thomson, R.I. 2008. A longitudinal study of ESL learners' fluency and comprehensibility development. *Applied Linguistics* 29: 359–380. doi: 10.1093/applin/amm041.

Derwing, T.M., Munro, M.J., and Wiebe, G.E. 1997. Pronunciation instruction for "fossilized" learners: Can it help? *Applied Language Learning* 8: 217–235.

Derwing, T.M., Munro, M.J., and Wiebe, G. 1998. Evidence in favor of a broad framework for pronunciation instruction. *Language Learning* 48: 393–410. doi: 10.1111/0023-8333.00047.

Derwing, T.M. and Rossiter, M.J. 2003. The effects of pronunciation instruction on the accuracy, fluency and complexity of L2 accented speech. *Applied Language Learning* 13: 1–18.

Derwing, T.M., Thomson, R.I., and Munro, M.J. 2006. English pronunciation and fluency development in Mandarin and Slavic speakers. *System* 34: 183–193. doi:10.1016/j.system.2006.01.005.

Derwing, T.M., Rossiter, M.J., Munro, M.J., and Thomson, R.I. 2004. Second language fluency: judgments on different tasks. *Language Learning* 54: 655–679. doi: 10.1017/S0272263109990015.

Derwing, T.M., Munro, M.J., Thomson, R.I., and Rossiter, M.J. (2009). The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition* 31: 533–557. doi: 10.1017/S0272263109990015.

Ellis, R. 2009. The differential effects of three types of task planningon the fluency, Ccomplexity, and accuracy in L2 oral production. *AppliedLinguistics* 30: 474–509. doi:10.1093/applin/amp042

Flege, J. E. 2009. Give input a chance! In T. Piske and M. Young- Scholten (Eds.) *Input Matters in SLA*, pp. 175–190, Bristol: Multilingual Matters.

Freed, B. 2000. Is fluency, like beauty, in the eyes (and ears) of the beholder? In: *Perspectives on Fluency*, H. Riggenbach (ed.), 243–265, Ann Arbor: MI: University of Michigan Press.

Gatbonton, E., and Segalowitz, N. 2005. Rethinking communicative language teaching: a focus on access to fluency. *The Canadian Modern Language Review* 61: 325–353. doi:.10.2307/3587026.

Guillot, M.-N. 1999. *Fluency and Its Teaching*, Clevedon: Multilingual Matters.

Hieke, A.E. 1981. A content-processing view of hesitation phenomena. *Language and Speech* 24(2): 147–160.

Holm, J. 2000. *An Introduction to Pidgins and Creoles*, Cambridge: Cambridge University Press.

Isaacs, T. and Thomson, R.I. (2013). Rater experience, rating scale length, and judgments of L2 pronunciation: revisiting research conventions. *Language Assessment Quarterly* 10(2): 135–159.

Kang, O. 2010. Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System* 38(2): 301–315. doi: 10.1016/j.system.2010.01.005.

Kormos, J. 2006. *Speech Production and Second Language Acquisition*, Mahwah, MJ: Lawrence Erlbaum Associates.

Kormos, J. and Dénes, M. 2004. Exploring measures and perceptions of fluency in the speech of second language learners. *System* 32(2): 145–164. doi: 10.1016/j.system.2004.01.001.

Larsen-Freeman, D. 1997. Chaos/complexity science and second language acquisition. *Applied Linguistics* 18(2): 141–165.

Larsen-Freeman, D. 2012. Complex, dynamic systems: a new transdisciplinary theme for applied linguistics? *Language Teaching* 46: 202–214 doi: 10.1017/S026144811000061.

Larsen-Freeman, D. and Cameron, L. 2008. *Complex Systems and Applied Linguistics*, Oxford: Oxford University Press.

Lee, J.G., Cadierno, T., Glass, W.R., and VanPatten, B. 1997. The effects of lexical and grammatical cues on processing past temporal reference in second language input. *Applied Language Learning* 8: 1–23.

Levelt, W. 1989. *Speaking: From Intention to Articulation*, Cambridge, MA: MIT Press.

Levelt, W. 1999. Producing spoken language: a blueprint of the speaker. In: *The Neurocognition of Language*, C. Brown and P. Hagoort (eds.), 83–122, Oxford: Oxford University Press.

Lowie, W. 2010. The development of early L2 phonology: a dynamic approach. In: *Proceedings of the International Symposium on the Acquisition of Second Language Speech, New Sounds*, vol. 6, Poznán, Poland. Retrieved November 15, 2013 from http://ifa.amu.edu.pl/newsounds/Proceedings_guidelines.

MacGregor, L.J., Corley, M., and Donaldson, D.I. 2009. Not all disfluencies are equal: the effects of disfluent repetitions on language comprehension. *Brain and Language* 111(1): 36–45.

Miller, G. 1956. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review* 63: 81–97. doi: 10.1037/h0043158.

Munro, M. J., and Derwing, T.M. 1995a. Foreign accent, comprehensibility and intelligibility in the speech of second language learners. *Language Learning* 45: 73–97. doi: 10.1111/j.1467-1770.1995.tb00963.x.

Munro, M.J., and Derwing, T. 1995b. Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech* 38,: 289–306.

Munro, M. J. and Derwing, T.M. 2006. The Functional Load principlein ESL pronunciation instruction: an exploratory study. *System* 34(4): 520–531. doi: 10.1016/j.system.2006.09.004.

Munro, M. J., and Derwing, T.M. 2008. Segmental acquisition inadult ESL learners: a longitudinal study of vowel production. *Language Learning* 58: 479–502. doi: 10.1111/j.1467-9922.2008.00448.x.

Munro, M.J., Derwing, T.M., and Morton, S. 2006. The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition* 28: 111–131. doi: 10.1017/S0272263106060049.

Nakuma, C. 1998. A new theoretical account of "fossilization": implications for L2 attrition research. *IRAL* 36(3): 247–257.

Nation, I.S.P. 1989. Improving speaking fluency. *System* 17: 377–384. doi: 10.1016/0346-251X(89)90010-9.

Reed, M., and Michaud, C. 2005. *Sound Concepts: An Integrated Pronunciation Course*, New York: McGraw-Hill.

Rossiter, M.J., Derwing, T.M., Manimtim, L., and Thomson, R.I. 2010. Fluency: the missing component in the "communicative" classroom. *Canadian Modern Language Review* 66: 583–606.

Saito, K. 2013. Re-examining effects of form-focused instruction onL2 pronunciation development: the role of explicit phonetic information.*Studies in Second Language Acquisition* 3: 1–29.

Saito, K., and Lyster, R. 2012a. Effects of form-focused instructionand corrective feedback on L2 pronunciation development of /r/ by Japaneselearners of English. *Language Learning* 62(2): 595–633. doi: 10.1111/j.1467-9922.2011.00639.x.

Saito, K., and Lyster, R. 2012b. Investigating pedagogical potentialof recasts for L2 vowel acquisition. *TESOL Quarterly* 46(2): 387–398. doi: 10.1002/tesq.25.

Schmidt, R. 1992. Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition* 14: 357–385. doi: 10.1017/S0272263100011189.

Schmidt, R. 2001. Attention. In: *Cognition and Second Language Instruction*, P. Robinson (ed.), 3–32, Cambridge: Cambridge University Press.

Segalowitz, N. 2010. *Cognitive Basis of Second Language Fluency*, New York: Routledge.

Selinker, L. 1972. Interlanguage. *IRAL* 10(3): 209–231.

Skehan, P. 1998. *A Cognitive Approach to Language Learning*, Oxford University Press.

Skehan, P. 2009. Modelling second language performance: integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics* 30: 510–532. doi: 10.1093/applin/amp047.

Skehan, P. and Foster, P. 2008. Complexity, accuracy, fluency and lexis in task-based performance: a meta-analysis of the Ealing research. In: *Complexity, Accuracy, and Fluency in Second Language Use, Learning, and Teaching*, S. Van Daele, A. Housen, F. Kuiken, M. Pierrard, and I. Vedder (eds.), Brussels: University of Brussels Press.

Thomson, R.I. 2011. Computer assisted pronunciation training: targeting second language vowel perception improves pronunciation. *CALICO Journal* 28: 744–765.

Thomson, R.I. 2012. Improving L2 listeners' perception of English vowels: a computer-mediated approach. *Language Learning* 62: 1231–1258. doi: 10.1111/j.1467-9922.2012.00724.x.

Thomson, R.I. and Isaacs, T. 2009. Within-category variation in L2English vowel learning. *Canadian Acoustics* 37(3): 138–139. Proceedings ofthe Annual Conference of the Canadian Acoustics Association, Niagara-on-the-Lake, ON.

Towell, R., Hawkins, R., and Bazergui, N. 1996. The development of fluency in advanced learners of French. *Applied Linguistics* 17(1): 84–119. doi: 10.1093/applin/17.1.84.

Trofimovich, P., and Isaacs, T. (2012). Disentangling accent from comprehensibility. *Bilinguialism: Language and Cognition* 15: 905–916. doi: 10.1017/S1366728912000168.

VanPatten, W. 1990. Attending to form and content in the input: an experiment in consciousness. *Studies in Second Language Acquisition* 12– 287–301. doi: 10.1017/S0272263100009177.

Walley, A.C. and Flege, J.E. 1999. Effect of lexical status on children's and adults' perception of native and non-native vowels. *Journal of Phonetics* 27(3): 307–332. doi: 10.1006/jpho.1999.0098.

Part IV Pronunciation of the Major Varieties of English

13 North American English

CHARLES BOBERG

"North American English" and "pronunciation": a definition of terms

When discussing varieties of English, many people identify the two dominant standard varieties as "British" and "American". This label is less than ideal, since what most people think of as "American English" is also spoken by a majority of Canadians, who do not consider themselves "American" in the normal sense of that word. As we will see, the English of most Canadians is actually closer to "General American English" than many of the regional and social types of English spoken in the United States. Especially in comparison with British or Southern Hemisphere varieties, Canadian English is incontestably a type of "American English", but in deference to the binational home of this type of English, the set of English varieties spoken on the North American continent will here be called "North American English" (NAE). One of these varieties, traditionally associated with parts of the midwestern and western United States and with central and western Canada, can now be heard, at least at higher social levels, across much of the continent. Beyond its native territory, it serves not only as a kind of pan-regional standard to be used in public domains like mass media communication and higher education, but as an acquisition target for learners of English as a second language and as a style-shifting target for many native speakers of other varieties of NAE, who wish to benefit from its high social prestige. This variety will be called "Standard North American English" (SNAE).

The term "pronunciation" in fact comprises many distinct types of sound difference. They are organized here into four levels of analysis. First, we will examine matters of phonemic contrast, or the "inventory" of phonemes in each variety. For instance, pairs of words that potentially differ by only one sound, such as *cot* and *caught*, or *bomb* and *balm*, are the same in some dialects but different in others, depending on whether the sounds they contain involve a phonemic

The Handbook of English Pronunciation, First Edition. Edited by Marnie Reed and John M. Levis.

contrast. Second, we will survey the phonological rules that cause systematic differences in the incidence of particular phonemes, usually involving context-dependent neutralizations of phonemic contrast. Third, we will identify examples of phonemic incidence that are lexically rather than phonologically conditioned; that is, rather than reflecting the operation of regular phonological rules that appeal to phonological categories, they are best understood as the unique properties of particular words. Fourth, we will describe differences in the phonetic quality of phonemes, such as the way the vowel of a word like *house* or *bad* or *stock* is pronounced in different regions of North America.

The first three types of variation we will call phonological. These entail contrast or alternation among phonemes, which we will represent in a broad transcription between forward slashes, indicating contrastive relations and historical word classes, rather than precise phonetic detail. The phonemic symbols used here will follow the binary tradition of American structuralism, as found throughout the work of Labov, in which the organization of English vowels into short and long subsystems, and the further division of long vowels into subsystems based on glide direction, is made explicit. The fourth type of variation we will call phonetic, which will be indicated in a narrow transcription, between square brackets. Such differences, which underlie the layman's concept of a regional "accent", are subphonemic and cannot be represented by the English spelling system. They will therefore require the phonetic precision of symbols from the International Phonetic Alphabet. We will also make use of the set of keywords developed by Wells (1982) to represent classes of English words that share the same historical vowel sound, with a common development from Middle English; these will appear in small capitals. Thus, the keyword DRESS represents the normal development of short /e/, as in words like *set*, *head*, *test*, *fell*, or *berry*, etc., while FACE represents the normal development of long /e/, or /ey/ (historically derived from Middle English long /ā/ via the Great Vowel Shift), as in words like *state*, *hay*, *paste*, *fail*, or *bare*, etc. The full set of broad phonemic symbols to be used in this chapter, with their equivalent keywords, is given in Table 13.1.

All of these aspects of variation and change have been well researched in a tradition that now reaches back almost a century in some areas. Regional variation in phonemic inventory and the phonetic quality of vowels were the main concern of the *Atlas of North American English* (Labov, Ash, and Boberg 2006, hereafter *ANAE*), which used auditory-impressionistic and acoustic phonetic analysis to examine a sample of approximately 700 participants from across the continent. The following discussion will often draw on data from this study, which provides the standard current treatment of these subjects, as well as from smaller studies on narrower topics. Regional variation in phonemic incidence is a major concern of an allied but older tradition of dialect research that extends back to the 1950s in Canada (studies of speech differences along the international boundary by Avis 1956 and Allen 1959) and to the 1930s in the United States (Kurath's dialect surveys of the eastern seaboard, which produced the summary treatment in *The Pronunciation of English in the Atlantic States* (Kurath and McDavid 1961), hereafter *PEAS*). Variation in phonemic

Table 13.1 Broad transcription of English vowel phonemes (Labov, Ash, and Boberg 2006) with keywords from Wells (1982).

Short/lax vowels (V)	*Long/tense vowels*			
	Front up-gliding (Vy)	*Back up-gliding (Vw)*	*Monophthongal/ in-gliding (Vh)*	*Pre-rhotic (-r)*
/i/ KIT	/iy/ FLEECE	/iw/ FEW, CUE	/æh/ BATH	/iyr/ NEAR
/e/ DRESS	/ey/ FACE	/uw/ GOOSE	/ah/ PALM,	/eyr/ SQUARE
/æ/ TRAP	/ay/ PRICE	/ow/ GOAT	/oh/ THOUGHT, CLOTH	/ahr/ START
/o/ LOT	/oy/ CHOICE	/aw/ MOUTH		/owr/ FORCE
/ʌ/ STRUT				/ohr/ NORTH
/u/ FOOT				/uwr/ CURE
				/ɜˇ/ NURSE

incidence is also exhaustively recorded, of course, by general-purpose dictionaries and by specialized dictionaries of pronunciation, like that of Kenyon and Knott (1953).

General pronunciation features of Standard North American English (SNAE): what makes people sound North American?

Phonemic inventory: how many phonemes occur in SNAE?

There is no need to review the inventory of SNAE consonant phonemes here: in most respects, this is identical with that of other varieties and is described elsewhere in this volume. Only one matter of phonemic contrast among consonants will be mentioned here: that involving the voiced and voiceless types of /w/, or /w/ and /hw/, as found in pairs like *wear* and *where*, *weather* and *whether*, *wine* and *whine*, or *witch* and *which*. While some conservative speakers of NAE maintain a distinction between these sounds, it has largely disappeared among younger speakers in most regions, so that it can be safely described as absent in SNAE.

Regional pronunciation differences in English are far more likely to involve vowels than consonants. In particular, there are four important variables of vowel contrast that distinguish major dialects of English, including SNAE. These are shown in Table 13.2.

The first line of Table 13.2 refers to the split of Middle English short /u/. This had affected southern English speech by the seventeenth century (Wells 1982: 197), early enough to be transplanted to North America, but never spread to northern

Table 13.2 Phonemic contrasts in the vowel systems of Standard British English (SBE) and North American English (SNAE). Parentheses indicate regional and/or social variation.

Contrast (≠)	*SBE*	*SNAE*
FOOT ≠ STRUT, or /u/ ≠ /ʌ/	YES	YES
TRAP ≠ BATH, or /æ/ ≠ /æh/	YES	NO
PALM ≠ LOT, or /ah/ ≠ /o/	YES	NO
LOT ≠ THOUGHT, or /o/ ≠ /oh/	YES	(NO)

England, where FOOT and STRUT still rhyme today. This variable therefore divides Britain into two dialect regions but unites the standard variety of British English, which is regionally rooted in the south-east of England, with SNAE, which has /u/ in FOOT but /ʌ/ in STRUT.

The remaining lines of Table 13.2 display important trans-Atlantic differences in phonemic inventory. The first involves another split, in this case of Middle English short /a/. Like the split of short /u/, this occurred in southern England, leaving the North unaffected. It seems to have occurred in two stages. First, in the seventeenth and early eighteenth centuries, short /a/ was lengthened to [a:] before voiceless fricatives and a few other environments (the BATH class), elsewhere remaining short and shifting forward to [æ] (the TRAP class; Wells 1982: 203–204). This aspect of the split did make it across the Atlantic, at least to some founding communities, though its subsequent history in American English is complicated and led to several of the dialect differences that will be discussed below. The second stage of the split was a backing of the lengthened vowel from [a:] to [ɑ:] in south-eastern England, producing the sound of the modern BATH class in Standard British English, with /ah/. This must have happened in the late eighteenth or early nineteenth century (Wells 1982: 234), early enough to be transplanted to Australia and New Zealand, but not to North America, where the lengthened vowel tended to be raised rather than backed. While some regions of North America retain a distinct, raised vowel in a much expanded version of the BATH class today, the split of short /a/ has collapsed in most NAE dialects. SNAE has a single phoneme, /æ/, in both TRAP and BATH words, with only subphonemic variation in phonetic quality (Kurath and McDavid 136; *ANAE* 173–174).

The last two lines of Table 13.2 involve mergers rather than splits. The first concerns the small remnant of Middle English long /ā/ that was not raised to /ey/ (FACE) in the Great English Vowel Shift, remaining in a low-central position. This includes the word *father*, along with lengthened /a/ before /-lm/ (*almond*, *alms*, *balm*, *calm*, *palm*, *psalm*) and a few other unusual words (*ma*, *pa*, *rah*, etc.). Its residual status made this PALM class prone to merger with neighboring vowels. In southern England it merged with the lengthened and retracted BATH class and, as a result of /r/ vocalization, with the START class. In North America, where BATH was not retracted and START generally retained its /r/, the tendency was instead

for PALM to merge with LOT, the regular development of Middle English short /o/. LOT began to shift down and forward from its original mid-back position by the seventeenth century, reaching a low-central unrounded position, approximately [ɑ], in some southern English dialects as North America's English-speaking colonies were being founded. Wells (1982: 245) suggests that it is not clear whether this happened in England or was an American innovation. It later moved back and re-rounded in standard British English, but the low-central unrounded vowel survives in much of North America, occupying the same phonetic range as southern British BATH. This accounts for the American tendency to transcribe LOT as /a/ or /ɑ/, based on its phonetic identity in many American dialects, rather than as /o/, based on its historical identity. In these dialects, LOT has merged with PALM, so that *father* rhymes with *bother* and *balm* and *bomb* are homophones (*PEAS* 141–142; *ANAE* 171). The main exception to this pattern is New England, as discussed below.

The last line of Table 13.2 also involves the contrastive relations of the LOT class, in this case with the THOUGHT class, sometimes referred to as long open /o/ and transcribed as /ɔ/ (the /oh/ symbol used here, like the /o/ for LOT, is consistent with the binary system of broad transcription mentioned above). Like PALM, THOUGHT is not the regular development of any single Middle English vowel phoneme, and its membership varies by dialect. For instance, as a result of a lengthening of Middle English /o/ before voiceless fricatives parallel to the development of Middle English /a/ described above, the THOUGHT class includes the CLOTH subclass in most American dialects but not in Standard British English, where CLOTH retains its original association with LOT. In North America, THOUGHT has therefore shown a similar tendency to merge with its neighbor, LOT, except where other phonetic developments have kept the two categories distinct (*ANAE* 123). These phonetic developments include some of the most salient regional differences in pronunciation, to be discussed below, but tend to occur in regions that are not the main source of SNAE (specifically, the Mid-Atlantic, the Inland North, and the South). Outside these areas, in SNAE, unrounding of THOUGHT and phonetic approximation to LOT has generally led to the "low-back merger", making homophones of such pairs as *cot* and *caught*, *sod* and *sawed*, *stock* and *stalk*, *don* and *dawn*, and *collar* and *caller*. This merger is now complete in northern New England, the West and Canada, as well as in parts of the Midland and South, (*ANAE* 170). It is in progress in the remaining parts of the Midland and South – more advanced in some communities and social groups than in others – and may even be making inroads among younger, upwardly mobile speakers in the areas that have historically resisted it, as the pronunciation features that prevented it in the past become socially stigmatized. Nevertheless, the retention of the LOT–THOUGHT contrast among many Americans who would consider themselves – and not without reason – to be speakers of SNAE compels us to place parentheses around the "NO" in the last line of Table 13.2, to concede the persistence of dialect variation in this respect.

To summarize this section, the distinctive sound of SNAE is strongly influenced by three important features connected with variables of phonemic inventory. Most North Americans use /æ/ in BATH words, so that they have the same vowel as

TRAP words; make no distinction between PALM and LOT words, because of the low-back to low-central, unrounded pronunciation of LOT; and, increasingly, also fail to distinguish LOT and THOUGHT words, using a lowered, unrounded vowel in THOUGHT that is too close to the vowel of LOT to support a stable phonemic contrast.

Systematic variation in phonemic incidence: words whose pronunciation varies by phonological rule

The most important and pervasive systematic variable of phonemic incidence in English is the occurrence of /r/ in "coda" position; that is, when it is not prevocalic, as in Wells' keywords NEAR, SQUARE, START, NORTH, FORCE, CURE, and NURSE. The tendency to delete or "vocalize" post-vocalic /r/ in English began in restricted environments in the Middle English period, but did not become a more general feature of British English until the eighteenth or nineteenth century, too late to be implanted with the initial English-speaking settlement of North America (Wells 1982: 218). Nevertheless, in the nineteenth century "*r*-lessness" became a defining feature of Standard British English, whence it spread across the Atlantic as a prestige feature to several regions along the east coast of the United States. These included eastern New England, New York City, and parts of the South, but not the intervening southern Mid-Atlantic region around Philadelphia, or Canada (*PEAS* 171, Map 156; *ANAE* 48). The original colonial dialects of NAE, already carried inland by westward migration, remained unaffected. In the late twentieth century, the prestige of *r*-less pronunciation remained high in Britain, where it continued to conquer new territory, but was reversed in the United States, in favor of the more "American" sound of a fully constricted post-vocalic /r/. This prestige reversal was examined by the most famous sociolinguistic study ever undertaken, in New York City, where /r/ was being re-inserted by middle-aged and younger speakers interested in upward social mobility in the 1960s (Labov 1966, 1972a).

Today, *r*-lessness remains a variable feature along the Atlantic seaboard, heard more from older speakers with local social networks than from the young or globally oriented. Its recession is more or less complete in European-American Southern speech, especially in large cities like Atlanta and Houston, but it has persisted to a greater extent in African-American speech, as mentioned below. It also continues to be widespread in eastern New England, even in large cities like Boston and Providence, and survives in popular culture in the catch phrase used as a stereotype of Boston speech, "pahk the car in Hahvuhd Yahd".

Even in communities that retain /r/ vocalization, however, it is usually variable. Its frequency responds both to stylistic factors, with less vocalization in formal, monitored speech, and to a range of phonological and other linguistic constraints. The most favorable environment for vocalization is the unstressed /ər/ of LETTER words, as in *letter*, *September*, *character*, or *spectacular*, but especially in word-internal position, as in *permission*, *Saturday*, *afternoon*, or *information*. This context seems to "fly below the radar" of speakers concerned with moving away from the

r-less speech they grew up with. Another vocalization context for otherwise /r/-constricting speakers is the presence of two *r*'s in a word, which promotes a kind of dissimilation in which one of them is vocalized or deleted. This applies to the first /r/ in words like *corner*, *former*, *ordinary*, and *quarter*, and to the second in words like *mirror* and *terror*, or when an agentive or comparative suffix <-er> is added to a stem ending in /r/, as in *bearer* or *fairer*. The most variable vocalization environments are the NEAR, SQUARE, START, NORTH, FORCE, and CURE sets that feature /r/ in their stressed syllables: these often have a restored /r/ in communities moving away from vocalization, but remain *r*-less in communities where this feature is more stable. Generally *r*-less speakers have the PALM vowel in the START class and the THOUGHT vowel in the NORTH and FORCE classes: Labov (1966, 1972a) investigates potential homophones like *dock* and *dark* and *sauce* and *source* in New York City. (Where NORTH and FORCE are different, as in traditional Boston speech, NORTH has [ɒ] and FORCE has an in-gliding variant of the GOAT vowel [ɔə].) The least favorable context for /r/ vocalization, and the first in which it is normally restored, is the syllabic /r/ or stressed /ɜ˞/ of NURSE words, like *her*, *girl*, *bird*, or *first*. Almost all Americans have a constricted /r/ in these words today, perhaps because a diphthongized variant of this vowel, pronounced as [ɜɪ], was the target of negative stereotypes of *r*-less dialects in the mid-twentieth century, as when New York cabbies were reputed to say "toity-toid street" for *thirty-third street*.

Most other systematic variables of phonemic incidence in English involve conditioned mergers, or neutralizations of phonemic contrast in particular phonological contexts. A whole set of these neutralizations is connected with the variation in /r/ just discussed, but involves inter-vocalic rather than post-vocalic /r/ (see *PEAS* 123–127; Gregg 1957b). It is a general property of English phonology that an /r/ in the coda of a syllable (post-vocalic /r/) limits the range of vowels that can occur before it. In particular, short vowels (/i, e, æ, o, ʌ, u/, or those of KIT, DRESS, TRAP, LOT, STRUT, and FOOT) are generally not licensed in this position. If we think of inter-vocalic /r/ in words like *very*, *carrot*, *orange*, and *hurry* as ambi-syllabic, at once closing the preceding syllable (coda position) and starting the next (onset position), we can see how the incidence of vowels in the first syllables of such words will be constrained by the variable presence of coda /r/. In *r*-less dialects, coda /r/ is not present, so that the full range of vowels can occur. In these dialects we therefore hear the vowel of DRESS in *very*, that of TRAP in *carrot*, that of LOT in *orange*, and that of STRUT in *hurry*. In the /r/-retaining dialects most commonly associated with SNAE, by contrast, we hear compromise vowels in these contexts that represent neutralizations of contrast between long and short vowels, which match the set of vowel qualities that occur before final coda /r/, in SQUARE, FORCE, and NURSE. The DRESS and TRAP vowels merge with FACE; the LOT vowel with GOAT; and the STRUT vowel with NURSE. Thus, *Mary*, *merry*, and *marry* all sound more or less like *merry*; *coral* sounds like *choral*; and *hurry* has the vowel of *her*.

Another common neutralization of vowel contrast affects the vowels /uw/ and /iw/ after the alveolar stops /t, d, n/. In most English dialects, this contrast has survived after labial and velar consonants (*boot* versus *beauty*; *coop* versus *cube*), where /iw/ is distinguished by a palatal glide before the vowel, [ju]. After liquids

and /s/, however, the contrast has now been lost, so that /uw/ now occurs in place of /iw/ in words like *pollution* and *super*. In NAE, this neutralization is generally extended to instances of /iw/ after /t, d, n/, as in *Tuesday*, *student*, *duty*, and *news*, though some conservative speakers retain a palatal glide in at least some of these words (*PEAS* 113, Map 33), especially in the American South (*ANAE* 55) and parts of Canada.

Another important systematic difference in phonemic incidence concerns the consonant /t/, which is partially voiced, or "flapped", when it occurs after a stressed vowel or vowel-liquid sequence and before an unstressed vowel or syllabic sonorant, as in *city*, *party*, *faulty*, *daughter*, or *battle*; flapping after /l/ is variable. Medial /d/ also has a slightly weakened articulation in these contexts, so that for many speakers the /t-d/ contrast is neutralized in pairs like *atom* and *Adam*; *coated* and *coded*; *rater* and *raider*; *metal* and *medal* (or *meddle*); *diluted* and *deluded*; etc. Flapping can also occur two syllables after the stress, as in *charity*, *monitor*, or *penalty*, but is more variable in this position. Following a stressed vowel and /n/, as in *twenty* or *winter*, the /t/ is often deleted altogether, so that *winter* and *winner* sound the same; preceding a syllabic /n/, as in *button* or *Latin*, it is replaced in most dialects with a glottal stop. Flapping in NAE is now completely standard, to the point where its absence, especially in the core environment after stressed vowels or vowel-/r/ sequences (*city* and *party*), is considered pompous or affected.

The foregoing discussion of systematic variation in phonemic incidence can be summarized by noting that SNAE retains coda or post-vocalic /r/, with a consequent reduction in the number of vowel contrasts before intervocalic /r/, so that *hairy*, *berry*, and *carry* all have the vowel of *berry*, and *forest* and *worry* have the vowels of *four* and *were* respectively; has also lost the contrast between /uw/ and /iw/ after /t, d, n/, so that *due* and *dew* sound like *do*; and replaces /t/ in post-tonic, inter-vocalic contexts with a sound that, for most speakers, is identical with /d/, so that *seated* and *seeded*, or *bitter* and *bidder*, are homophones.

Lexical variation in phonemic incidence: words whose pronunciation varies in phonologically irregular ways

Some variables of phonemic incidence are truly idiosyncratic. A classic example from American dialectology is the fricative in the word *greasy*, which varies between /s/ in the North and /z/ in the Midland and South of the Eastern United States (*PEAS* 176–177, Map 171). The irregular nature of this variation is demonstrated by its absence in phonologically similar words like *easy*, *teasing*, *fleecy*, or *increasing*: variation between /s/ and /z/ is clearly a property of the word *greasy*, not a phonological rule affecting intervocalic /s/. The same could be said about the word *vase*, which varies in both its vowel and final consonant, rhyming alternately with *face*, *phase*, or *spas* (PEAS 177); this pattern is not observed in similar words like *base* or *raise*, etc. A lack of systematicity has made this type of variation less interesting to phonologists, but not to the general public. Many of the most frequently cited examples of dialect variation involve pronunciations

of particular words. Further examples from *PEAS* include the vowels of DRESS, KIT, or FACE in *again* (131); of TRAP, PALM or LOT in *aunt* (135); of DRESS or TRAP in *catch* (139); of FLEECE or KIT in *creek* (148); of FLEECE or PRICE in *either* and *neither* (149); of GOOSE or FOOT in *roof* (154) and *root* (155); or of LOT or THOUGHT in several words, including *sausage*, *water*, *wash*, *fog*, *long*, and *on* (162–164); or the voicing of the fricative in *without* (176). Nevertheless, though such words may excite the interest of casual observers, it is difficult to give a general account of them, since each tends to display a unique regional distribution.

Somewhat less idiosyncratic are several sets of words of varying size that display more or less regular differences between SNAE and Standard British English. One of them, already mentioned above, is the CLOTH set: words that, in British English, feature short /o/ before voiceless fricatives, as in *coffee*, *lost*, or *boss*. In North American dialects that distinguish LOT and THOUGHT, most of these words have the vowel of THOUGHT (though there are exceptions, generally involving less frequent words). The same is true of most words that feature this vowel before /g/ or /ŋ/, like *dog*, *log*, *song*, and *wrong*, though phonemic incidence in this subset is even more variable, as mentioned above. For most people, *cog* and *gong*, for example, have the LOT vowel, as does the new word *blog*, while *hog* varies by region.

Another loosely cohesive set of British English–NAE differences involves reduction or deletion of unstressed vowels, for instance in the set of Latinate words that end in *-tary* or *-tory*, such as *secretary*, *military*, *preparatory*, or *mandatory*. The penultimate vowel of these words is usually deleted or reduced in Britain but preserved in North America. NAE also retains unstressed vowels in words like *medicine*, *police*, and the names of berries (*blackberry*, *raspberry*, *strawberry*, etc.), as well as in place names like *Birmingham* and *Manchester* and names beginning with *Saint*. When we add to this the evidence of other distinctively British shortenings in words like *forehead* and *waistcoat*, NAE appears the more generally conservative dialect in this respect.

Within North America, some of the differences between Canadian and American English also involve variation in phonemic incidence, with variable adherence to British norms in Canada. Several of these involve the PRICE vowel, /ay/. Words in *-ile*, such as *fertile*, *futile*, *hostile*, *missile*, *mobile*, and *sterile*, have a reduced vowel in the second syllable, like that of *noble*, in American English but a full PRICE vowel, like that of *profile*, in Britain and Canada (Avis 1956: 46). On the other hand, Americans tend to have /ay/ in *-ine* words, like *genuine*, and in the Latin prefixes *anti-*, *semi-*, and *multi-*, where /i/ (KIT) is heard for the first set and /iy/ (FLEECE) in the second in Britain and, variably, in Canada (Avis 1956: 47). Verbs that begin with the prefix *di-*, like *digress*, *direct*, *dissect*, and *diverge*, along with their nominal forms, also vary between the PRICE and KIT vowels on either side of the Atlantic as well as within NAE, though in the case of *vitamin*, North Americans are united in using /ay/ in contrast to British /i/. Similarly irregular is a set of words that contain the Latin prefix *pro-*, such as the nouns *process*, *produce*, and *progress*: these tend to have the LOT vowel in the United States but vary between LOT and GOAT in Canada (Avis 1956: 45).

The cases of lexically governed phonemic incidence discussed so far have involved sets of a few dozen words at most, but there is one case that involves not dozens but hundreds or thousands of words. This is the set of "foreign (a)" words, discussed in Boberg (2010: 137–140) and first studied systematically in earlier work cited therein: words borrowed from other languages in which the stressed vowel is spelled with the letter <a>. Because English <a> has several phonemic values, these words can have the vowel of FACE, like *potato*, of TRAP, like *tobacco*, or of PALM, like *spa*. Most recent borrowings get either TRAP or PALM, but national dialects of English have different systems for deciding which vowel goes in which words. British English bases its choice mostly on vowel length. Since /æ/ is a short vowel, it is used before voiceless consonants, which tend to shorten preceding vowels, especially when they are spelled with double letters. Thus, /æ/ is heard in British English in *macho*, *mafia*, *pasta* and *Picasso*; exceptions are *taco*, which is variable, perhaps due to American influence, and *Iraq*, which usually has /ah/. A following voiced consonant encourages the preceding vowel to be treated as long: long /ah/ is preferred in the stressed syllables of *avocado*, *Colorado*, *drama*, *façade*, *lager*, *lava*, *llama*, *pajamas*, *Pakistani*, *panorama*, *plaza*, *Slavic*, and *soprano*; one exception to this is *lasagna* (or *lasagne*), which usually has /æ/. In American English, vowel length is far less important than the foreign status of the words, which demands the use of /ah/ rather than /æ/, perhaps on the model of Spanish, the most familiar "foreign" language in many parts of the United States. All of the foreign (a) words just listed have /ah/ in American English, except *Pakistani*, *panorama*, and *soprano*, which normally have /æ/, and *Colorado*, *Iraq*, and *pajamas*, which vary between /ah/ and /æ/. Canadians have a third pattern all their own, in which most of these words, at least traditionally, had /æ/, though some (like *façade*, *lasagna*, *lager*, *macho*, and *mafia*) have now begun to switch over to /ah/, apparently under American influence. Even most younger Canadians today, however, continue to use /æ/ in *avocado*, *Colorado*, *drama*, *Iraq*, *lava*, *pajamas*, *Pakistani*, *panorama*, *pasta*, *Picasso*, *plaza*, *Slavic*, and *soprano*, a list that includes several words (*avocado*, *drama*, *lava*, and *Slavic*) in which both Britons and Americans agree on /ah/. The Canadian preference for /æ/ as the default vowel for these words likely has its origins in the conventional Canadian understanding that where Britons say /ah/ in BATH words, Canadians say /æ/; this correspondence was simply transferred to foreign words as well, so that if British English had /ah/ in *avocado*, *drama*, *lava*, or *Slavic*, this should be rendered as /æ/ in Canada.

Phonetic realization: what is the phonetic quality of each phoneme?

Even variables of phonemic incidence that involve large sets of words, like the foreign (a) class just discussed, are limited in their role as indicators of dialect difference by their frequency of occurrence: while speakers may react strongly to unfamiliar or different pronunciations of words like *vase*, *roof*, *greasy*, or *pasta* when they hear them, the likelihood of any one of these words occurring in ordinary discourse is fairly small. Far more likely is that any substantial quantity of speech will include several examples of the more common vowel phonemes. Given their high frequency in discourse, as well as their systematic and regular application, variables of phonetic quality must therefore

Table 13.3 Approximate phonetic quality of the 14 vowel phonemes of Standard North American English, including pre-rhotic variants.

	Long/tense vowels			
Short/lax vowels	*Front up-gliding*	*Back up-gliding*	*Monophthongal*	*Pre-rhotic*
/i/ KIT [ɪ]	/iy/ FLEECE [ɪi]	/iw/ FEW, CUE [jɪu]	/o-ah-oh/ LOT, PALM, THOUGHT, CLOTH [ɑː, ɒː]	/iyr/ NEAR [ɪɹ]
/e/ DRESS [ɛ]	/ey/ FACE [ɛɪ]	/uw/ GOOSE [ʉu, ɨu]		/eyr/ SQUARE [ɛɹ]
/æ/ TRAP, BATH [æ]	/ay/ PRICE [aɪ]	/ow/ GOAT [ɵʊ]		/ahr/ START [ɑɹ]
/ʌ/ STRUT [ʌ]	/oy/ CHOICE [ɔɪ]	/aw/ MOUTH [ɑʊ]		/owr/ NORTH, FORCE [ɔɹ]
/u/ FOOT [ʊ]				/uwr/ CURE [ʊɹ, ɜˇ]
				/ɜˇ/ NURSE [ɜˇ]

play a leading role in allowing speakers to project their own regional or social identities, as well as to perceive and assess the identities of others whose speech they hear.

There is a great deal of regional variation in the phonetic value of phonemes across North America, as recorded by the *ANAE*, or by Thomas (2001). Analysis of this variation will be reserved for the next section. Here, in Table 13.3, we offer a general statement of the approximate phonetic quality of the vowels of SNAE. For other, analogous descriptions, see, *inter alia*, Wells (1982: 121–122), Kretzschmar (2004: 263–264) or Ladefoged (2006: 39); an earlier equivalent appears in Bloomfield (1933: 91). Where substantial inter-speaker variation occurs even within SNAE, two phonetic symbols appear. Allophonic variation due to phonetic context is assumed rather than explicitly indicated, so that the values in Table 13.3 indicate the main quality of each vowel, rather than the total range of its allophones.

Regional variation in NAE pronunciation

The most important regional differences in the pronunciation of NAE – variation in the phonetic qualities listed in Table 13.3 – arise from underlying differences in the set of phonemic contrasts portrayed in Table 13.2. The pronunciation of vowels, as observed by Martinet (1955), is governed by an "economy" of contrastive

relations in a limited vowel space. Each phoneme occupies a field of dispersion within this space and requires a surrounding margin of security – a kind of buffer zone – that keeps it distinct from neighboring phonemes. Normally, vowels make maximal use of the available space by arranging themselves evenly and symmetrically across it. Distinctions (the maintenance of contrast between neighboring phonemes) take up more space than mergers (the loss of contrast); the contrastive relations of a vowel therefore affect its available space and its phonetic quality. In addition, especially in complex vowel systems like that of English, phonetic or sociolinguistic forces occasionally produce a shift in the quality of a vowel, so that it begins moving through the vowel space. If its movement encroaches on a neighboring vowel, two developments are possible: a merger, which tends to limit further changes by creating extra space for the remaining phonemic distinctions, thus relieving pressure on surrounding vowels; or a chain shift, in which the shifting vowel causes responsive shifts in neighboring vowels, until a stasis is re-established, either by a new arrangement of the vowels or by a merger.

This theory of vowel systems motivates the analysis of dialect differences in the *ANAE*, like those of its predecessors, Labov, Yaeger, and Steiner (1972) and Labov (1991). Its overall view of NAE dialects comprises as many as 20 regional divisions (*ANAE*: 146, 148), depending what qualifies as a region or dialect, but the organization of chapters in its section on regional patterns suggests seven major regions, some of which contain important subdivisions: the North, Canada, New England, the Mid-Atlantic, the South, the Midland, and the West; these are shown in Figure 13.1, reproduced from *ANAE* Map 11.15 (148). Here, a broadly similar taxonomy is adopted, reflected in the titles of the following subsections. Matters of phonemic contrast and associated vowel shifts will be discussed under the appropriate regional subtitles. Labov places particular importance upon the phonemic status of the low-front and low-back corners of the vowel space, which he labels "pivot points" (Labov 1991: 12), given their crucial influence on regional phonetic patterns. The initial question to ask about any regional dialect of NAE is whether TRAP and BATH, in the low-front quadrant, and LOT and THOUGHT, in the low-back quadrant, are one phoneme or two.

New England: Boston and Providence

Though New England is often thought of as a unified region in historical and cultural terms, it embraces several distinct dialect areas (*PEAS*, Map 2; Boberg 2001). The LOT-THOUGHT variable divides it into a northern half, from Vermont to Maine, including Boston, where this distinction has been lost, and a southern half, from Connecticut to Rhode Island, including Providence, where it is maintained. Bisecting this division is a line separating eastern New England, including Boston and Providence, which is traditionally *r*-less, from western New England, including Springfield and Hartford, which remained *r*-full. Northeastern New England, including Boston, traditionally resisted the merger of PALM and LOT, which has affected the rest of North America, because LOT was merged instead

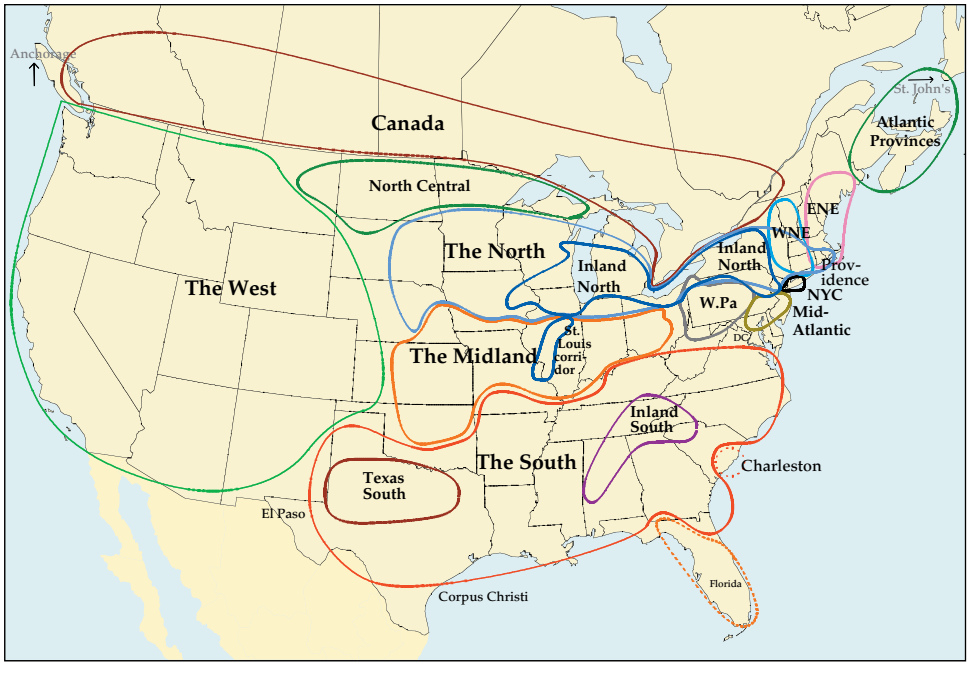


Figure 13.1 Map 11.15 from *The Atlas of North American English* (Labov, Ash, and Boberg 2006). Reproduced by kind permission of Mouton de Gruyter.

Table 13.4 Vowel qualities in traditional Boston English.

Vowel	*Quality*	*Vowel*	*Quality*
/eyr/ SQUARE	[ɛə]	/ohr/ NORTH	[ɒ]
/æ/ TRAP	[æ]	/owr/ FORCE	[ɔə]
/æh, ah, ahr/ BATH, PALM, START	[aː]	/o, oh/ LOT, THOUGHT	[ɒː]

with THOUGHT. In traditional Boston speech, some members of the BATH class were identified with the PALM class, as in British English, rather than with TRAP, though this pattern is now recessive. Northeastern New England also held out against the merger of NORTH and FORCE found in SNAE, but this, too, is fading with time (Laferrière 1979). The quality of the back up-gliding vowels of GOOSE, GOAT, and MOUTH tends to be conservative across New England, with less centralization and fronting than occurs farther south. Table 13.4 lists some vowel qualities typical of traditional Boston speech, which can be compared with those given for SNAE in Table 13.3.

The Mid-Atlantic: New York City, Philadelphia, and Baltimore

The Mid-Atlantic region between New England and the South is also bisected by the /r/ line, with New York City and region, to the north, being traditionally *r*-less and Philadelphia and Baltimore, to the south, being the major exception to this pattern along the east coast. In terms of Labov's pivot points, however, the Mid-Atlantic region is more unified than New England: its northern and southern sections share a common vowel system in which phonemic distinctions are maintained in both corners of the vowel space, with minor differences in lexical distribution. The low-back merger has been resisted by shifting THOUGHT (including the CLOTH subset) up to mid-back position, where it becomes a diphthong with a central in-glide, in the range between [ɔə] and [ʊə]. Here it is easily distinguished from the [ɑ] of LOT, but merges, in New York, with NORTH/FORCE (*sauce* = *source*). Labov (1966) showed that the height of the diphthong nucleus was an important sociolinguistic variable in New York, with higher qualities receiving a negative evaluation even from New Yorkers themselves. In the low-front quadrant, the TRAP-BATH split displays a parallel development: an upward shift of the tense vowel, BATH, along the front periphery of the vowel space. Its quality ranges from [ɛə] to [ɪə], with a parallel social evaluation to that of raised THOUGHT, and a parallel tendency to merge with SQUARE in New York (*bad* = *bared*). TRAP remains in the low-front position, at [æ]. As in most of North America, PALM is merged with LOT; /r/ vocalization in New York adds the START class to this set (*dark* = *dock*). These developments are summarized in Table 13.5.

One way of distinguishing New York City and Philadelphia, apart from vocalization of /r/, is in the lexical distribution of tense and lax vowels (*ANAE* 173). While we have used Wells' keywords to represent the tense vowels, the membership of these sets in NAE dialects is larger than in British English, for which the keywords were designed. The split of Middle English short /a/ in the Mid-Atlantic region, in particular, has received a great deal of scholarly attention because of the complexity of the conditioning factors that determine which vowels are tense, like BATH (e.g., Labov 1972a: 72–75; *ANAE* 173). In New York, the tensing environment was extended from the British BATH set, before voiceless fricatives, to

Table 13.5 Vowel qualities in traditional New York City English.

Vowel	*Quality*	*Vowel*	*Quality*
/eyr/ SQUARE	[ɛə]	/ohr, owr/ NORTH, FORCE	[ɔə]
/æh/ BATH	[eə, ɛə]	/oh/ THOUGHT, CLOTH	[ɔə, ʊə]
/æ/ TRAP	[æ]	/ah, ahr, o/ PALM, START, LOT	[ɑː]

vowels before voiced stops (*cab*, *bad*, *badge*, and *bag*) and front nasals (*ham*, *band*), though several nonphonetic constraints create exceptions to this rule. Philadelphia has tensing in a smaller range of environments and with more exceptions: among voiced stops, only /d/ causes tensing and the single word *sad* is a notable exception even to this. In the back vowels, the distribution of words like *chocolate*, *laundry*, and *sausage* between the /o/ and /oh/ or LOT and THOUGHT classes also shows regional variation. Particularly noteworthy is the preposition *on*, which rhymes with *don* in New York (as in the North generally) but with *dawn* in Philadelphia (as in the Midland and South; *ANAE* 189).

The Inland North: Chicago, Detroit, Cleveland, and Buffalo

The Inland North extends along the southern shores of the Great Lakes, from Milwaukee, Wisconsin, through Chicago, Detroit, Toledo, Cleveland, and Buffalo to Rochester, New York. Here, TRAP and BATH are a single phoneme, which has undergone the same phonetic development as BATH in the Mid-Atlantic territory, raising to mid-front position, approximately [ɛɐ] or [eə]. This has left room in the low-front quadrant, still occupied by TRAP in the Mid-Atlantic vowel system, for the forward shift of LOT (merged with PALM) to [a], which maintains its contrast with THOUGHT along a front-back dimension. The raising of TRAP-BATH and fronting of LOT-PALM are the initial and most striking components of the Northern Cities Vowel Shift (NCS, Labov 1991: 15–17; *ANAE* 187–191), which also involves several consequent developments in the short vowel subsystem. Fronting of LOT-PALM allows THOUGHT to unround and move down to [ɑ], which in turn makes room for STRUT to approach the lower-mid back region of the vowel space; this frees up the lower-mid central space, into which DRESS is retracted; finally, KIT is lowered into the lower-mid front space once occupied by DRESS. The *ANAE* uses the NCS to define its Inland Northern dialect region (204), though recent research by McCarthy (2011) suggests that its initiating stages, the raising of /æ/ and the fronting of /o-ah/, are no longer active changes, at least in her sample of Chicago speakers. The effects of the NCS are illustrated in Table 13.6, though it should be noted that it represents extreme targets for vowel shifting that are not reached by all speakers or in all contexts; they are intended to indicate the direction of shift.

The Inland North was initially settled mainly from New England, with which it shares several general northern characteristics, such as the conservative treatment of GOOSE, GOAT, and MOUTH referred to above, with relatively little movement

Table 13.6 Vowel qualities in Inland Northern speech (fully shifted).

Vowel	*Quality*	*Vowel*	*Quality*
/i/ KIT	[ə]	/ʌ/ STRUT	[ɔ]
/e/ DRESS	[ʌ]	/oh/ THOUGHT, CLOTH	[ɑː]
/æ, æh/ TRAP-BATH	[eə]	/ah, o/ PALM, LOT	[aː]

away from the rear periphery of the vowel space. As a result, MOUTH [ɑʊ] is articulated further back in the Inland North than PRICE [aɪ], the opposite of what we find farther south.

The South: Richmond, Charlotte, Atlanta, Nashville, Dallas, and Houston

In the Southern United States, an entirely different set of vowel shifts, known as the Southern Shift (Labov 1991: 25; *ANAE* 242–254), has developed in response to the most frequently cited element of Southern phonology, the monopthongization of /ay/, the vowel of PRICE. (In this case Wells' choice of keyword is unfortunate, since many Southerners do not monopthongize /ay/ before voiceless consonants; glide deletion happens most frequently in open syllables and before voiced consonants, so that *prize* would be a more appropriate keyword.) The realization of /ay/ as [a:] created a hole at the bottom of the subsystem of front up-gliding vowels that has pulled the nucleus of /ey/ (FACE) down toward low-central position, [ɐɪ], with /iy/ (FLEECE) following it downward in a chain shift. In their shifted positions, these long vowels have switched places with their short counterparts, KIT and DRESS, which have become tense, inward-gliding diphthongs, with nuclei higher and fronter than those of the originally long vowels. The third short front vowel, TRAP-BATH, is also tensed and diphthongal, or even triphthongal, with an upward then downward contour, especially in the BATH subset, which may parallel the lengthening of this class in other dialects.

Labov's description of the Southern Shift also includes a second component involving a parallel forward shift of the long back-upgliding vowels of GOOSE, GOAT, and MOUTH (Labov 1991: 25). The last of these is part of the Southern strategy for avoiding the low-back merger: as MOUTH shifts forward from [ɑʊ] to [æʊ], THOUGHT develops a back-upglide, so that it adopts the [ɑʊ] quality that MOUTH has in the North. It is thereby differentiated from LOT, which remains in low-back position and is lengthened but monophthongal, with a nuclear quality that often overlaps that of THOUGHT (*ANAE* 127, 254). Two short back vowels, /u/ (FOOT) and /ʌ/ (STRUT), are also lengthened and strongly centralized in Southern speech, while the up-glide of /oy/ (CHOICE) is either shortened, as in *boy* ([bɔə]), or deleted, as in *boil* ([bɔ:l]). Together, these ten vowel shifts combine to create what is known in popular culture as a "Southern drawl": their effects are summarized in Table 13.7.

The full set of shifts shown in Table 13.7 is only found among some speakers in the South and in some subregions more that others. Monophthongization of /ay/, which appears to be the initiating development of the Southern Shift, displays the widest spatial distribution, being found over most of what most people consider to be the South in a broader cultural sense: from Texas to Virginia, and from Kentucky, on the Ohio River, down to Mississippi and Alabama, on the Gulf of Mexico (*ANAE* 131). Glide deletion before a smaller class of liquid and nasal consonants, in words like *tile*, *tire*, and *time*, is variably found in an even larger region, reaching across the Ohio River into parts of the southern Midland, from

Table 13.7 Vowel qualities in traditional Southern speech.

Vowel	*Quality*	*Vowel*	*Quality*
/iy/ FLEECE	[əi]	/uw/ GOOSE	[ɨu]
/i/ KIT	[iə]	/u/ FOOT	[ɜː]
/ey/ FACE	[ɐɪ]	/ow/ GOAT	[ɜʊ]
/e/ DRESS	[eə]	/ʌ/ STRUT	[ɐː]
/æ, æh/ TRAP, BATH	[ɛɐ, aɪə]	/oy/ CHOICE	[ɔə]
/ay/ PRICE	[aː]	/o/ LOT	[ɒː]
/aw/ MOUTH	[æʊ, aɪʌ]	/oh/ THOUGHT	[ɒʊ]

southern Illinois across to Philadelphia; the *ANAE* therefore specifies glide deletion before obstruents and word-finally, as in *tide* and *tie*, as the diagnostic criterion for the South. Within the region established by this criterion, there are two subregions where the Southern Shift is particularly advanced, both in inland rather than coastal areas: one in North Texas, from Lubbock to Dallas; the other in the Appalachian region, including eastern Tennessee, western North and South Carolina, and northern Georgia and Alabama.

In the remainder of the South, including older coastal enclaves like Ocracoke Island, Charleston, Savannah, and New Orleans, the Southern Shift is less consistently present, displaying only a subset of its elements, or sociolinguistic variation within communities (for Ocracoke, see Wolfram and Schilling-Estes 1997; for Charleston, see Baranowski 2007). It is almost entirely absent from areas subject to strong non-Southern influence, including Washington, DC, on the northern edge of the South, and central and southern Florida: Orlando, Tampa, and Miami are not Southern cities in the linguistic sense. Moreover, unlike the Northern Cities Shift, which is most advanced in the major cities of the Great Lakes region, the Southern Shift is associated instead with smaller towns and rural areas, where it is identified with traditional Southern culture: Thomas (1997) documents this urban–rural split in Texas. Younger speakers in the largest urban centers of the South – especially Dallas, Houston, and Atlanta – often lack most or all components of the Southern Shift, which appears to be receding over time (*ANAE* 253). While back vowel fronting remains a vigorous change, supported by parallel developments in other regions, the front part of the Southern Shift is now subject to negative social evaluation and therefore rejected by young, urban speakers, particularly women; Fridland (2001) reports this development in Memphis, Tennessee, and Dodsworth and Kohn (2012) confirm it in Raleigh, North Carolina.

The Midland: Pittsburgh, Columbus, Cincinnati, Indianapolis and St. Louis

The existence and geographic extent of a Midland dialect region has been the subject of considerable debate among students of American dialect. Kurath proposed a broad Midland region between his North and South, extending westward from

Philadelphia into the Appalachian Mountains and southern Midwest (*PEAS*, Map 2). Subsequent analyses, including that of the *ANAE*, have treated the Midland as a transition zone, characterized more by a gradual recession of Southern features as one moves north than by unique features of its own. Thomas (2010) examines the North–Midland boundary in Ohio; Habick (1993) reports on aspects of the Southern Shift heard in central Illinois; and Marckwardt (1957) and Frazer (1978) show the transitional nature of Midland speech across the entire North–Central region. The *ANAE* finds strong fronting of back up-gliding vowels across the Midland, as well as a tendency to merge PALM-LOT and THOUGHT, already complete in the Pittsburgh area. Unlike the Inland North, whose cities display a uniform development of the NCS, Midland cities, like Pittsburgh (Johnstone and Kiesling 2008), Cincinnati (Boberg and Strassel 2000), and St. Louis (Murray 1986), are characterized by somewhat greater diversity. Pittsburgh, for instance, displays a monophthongization of /aw/ (*downtown* stereotypically becomes *dahntahn*); Cincinnati has its own, simpler version of the Mid-Atlantic tensing and raising of /æh/ (BATH); while St. Louis has a unique system of back vowels before /r/, with NORTH distinguished from FORCE but merged with START (*horse* and *hoarse* are different, but *born* and *barn* are the same).

The West: Denver, Phoenix, Seattle, San Francisco, and Los Angeles

Little can be said about the West beyond what was said about SNAE above: there is almost nothing to distinguish them. The double merger of PALM, LOT, and THOUGHT is complete throughout the region: Reed (1952: 186–187) reported its progress in Washington State two generations ago. The West also has a single /æ/ vowel with raising only before nasals (in *band* and *ham*) and a more moderate fronting of back up-gliding vowels than is found in the Midland or South. Along the West's eastern edge, several cities have a transitional status between the West and other regions. The largest of these is Minneapolis-St. Paul, which has the low-back merger of the West but the general raising of a unified TRAP-BATH vowel characteristic of the Inland North, as well as a typically northern resistance to the shifting of long up-gliding vowels, so that FACE and GOAT have an almost monophthongal quality, [e:] and [o:]. Much of the Great Plains region, including cities like Des Moines, Omaha, Kansas City, and Tulsa, displays a mixture of Midland and Western features, to the extent that these can be distinguished.

Canada: Vancouver, Edmonton, Calgary, Toronto, Ottawa, and Montreal

Like the West, most of Canada features a type of English that is difficult to distinguish from SNAE. The double low-back merger of PALM, LOT, and THOUGHT is complete across the country and TRAP and BATH are a single phoneme, /æ/, with raising only before nasals (Boberg 2010: 125–130). Canada shares the fronting of

GOOSE with much of the United States but has comparatively little centralization of GOAT (Boberg 2010: 144). Nevertheless, two phonetic variables do distinguish Canadian English from neighboring American dialects, in addition to the retentions of British phonemic incidence and a unique foreign (a) pattern, discussed above. In Ontario, the most important of these is the Canadian Shift, a vowel shift that involves an opposite development of TRAP and LOT to that found across the border in the American Inland North (Boberg 2000). First identified by Clarke, Elms, and Youssef (1995) and confirmed as a change in progress by later work (*ANAE* 216–224; Boberg 2010: 230), the Canadian Shift involves a retraction of TRAP into the low-central position left empty by the low-back merger. As TRAP moves back, DRESS moves down toward the low-front quadrant, pulling KIT down behind it. Retracted Canadian TRAP in Ontario has the same phonetic quality as fronted LOT across the border in south-eastern Michigan and western New York: a Detroit or Buffalo pronunciation of *solid* might be misunderstood as *salad* in Toronto, and a Toronto pronunciation of *black* might be mistaken for *block* in Detroit or Buffalo.

The stark cross-border difference found around the Great Lakes gradually weakens as one moves west, until it all but disappears on the Prairies and the Pacific coast. Here, a common low-back merger prevents the Canadian Shift from being as distinctive as it is further east; Kennedy and Grama (2102), in fact, report a similar development in California. Instead, another feature, Canadian Raising, provides a more subtle degree of difference. First described in Ontario English by Joos (1942) and in Vancouver by Gregg (1957a), and later studied more extensively by Chambers (1973) and Boberg (2010: 149–151; 204–205), Canadian Raising produces non-low nuclei in the low diphthongs /aw/ and /ay/ (MOUTH and PRICE) before voiceless obstruents. Thus, *cow*, *proud*, *tie*, and *tide* have low nuclei, [aʊ] and [aɪ], whereas the nuclei of *house*, *doubt*, *tight*, and *spice* are raised to lower-mid position, ranging from [ɛʊ] to [ʌʊ] for /aw/ and from [ɐɪ] to [ɜɪ] for /ay/. Raising of /ay/ in pre-voiceless environments has also been noted in a number of American dialects: most famously on Martha's Vineyard, Massachusetts, by Labov (1963), but also across much of the northern United States (*ANAE* 205–206). Raising of /aw/, by contrast, is more uniquely Canadian, and has inspired the most common American stereotype of Canadian speech, which has Canadians saying *oot and a boot* for *out and about* (like most stereotypes, this one is an exaggeration; in Western Canada, where the raised vowel is further back than in Ontario, a more accurate re-spelling would be *oat and a boat*). Boberg (2010: 156) demonstrates that a combination of Canadian Raising of /aw/ and retraction of /æ/ in the Canadian Shift separates most young speakers of Canadian English from their American peers, some of whom display moderate versions of one or the other feature but not both. The phonetic effects of the Canadian Shift and Canadian Raising are indicated in Table 13.8.

Boberg (2010) finds that the type of Canadian speech portrayed in Table 13.8 is particularly dominant across western and central Canada, from British Columbia to Ontario; it is also heard among some ethnic groups, particularly British-Canadians, in Montreal and to an increasing extent in Atlantic Canada, especially

Table 13.8 Vowel qualities in Canadian speech.

Vowel	*Quality*	*Vowel*	*Quality*
/i/ KIT	[ɐ̟]	/ay/ PRICE	[ɜɪ]
/e/ DRESS	[ɐ̝, ʌ]	/aw/ MOUTH	[ɜʊ, ʌʊ]
/æ, æh/ TRAP, BATH	[aː]	/ah, o, oh/ PALM, LOT, THOUGHT	[ɒː]

among younger, upwardly mobile people. Older, more locally oriented people in eastern Canada tend to speak a wider variety of local dialects, which limited space prevents us from discussing here: from traditional enclaves in the Ottawa Valley of eastern Ontario and several parts of the Maritime provinces to the highly distinctive dialects of Newfoundland, established by early nineteenth century settlement from south-western England and south-eastern Ireland (see Clarke (2004, 2010) for a description of Newfoundland pronunciation, which includes a low-central, unrounded vowel for LOT and a mid-back, rounded vowel in STRUT, in contrast to their usual qualities in mainland Canadian English).

Social variation in NAE

While the main focus of this chapter has been on regional differences in NAE, social differences also play an important role. There is no space here to discuss these in any detail, but the most obvious social divisions arise from ethnic differences (Boberg 2012), since socioeconomic differences *per se* tend to be reflected more in grammatical than in phonological variation. Most large American cities now feature three main ethnic dialects. The regional types described above are associated mostly with the population of European ethnic origin. African-Americans, though a diverse group themselves, tend to participate less in local European-American speech patterns, particularly at lower social levels. Most of them maintain instead a basically Southern type of speech, reflecting the migration of large numbers of African-Americans from the South to northern and western cities, from the late nineteenth to the mid twentieth century. African-American English (AAE) has been extensively studied (e.g., Wolfram 1969; Labov 1972b) and no attempt will be made here to review that body of research. Suffice it to say that the most distinctive aspects of AAE involve grammatical rather than phonological variables; among the latter, the most distinctive in most American cities is the vocalization of /r/, which has become an ethnolinguistic variable in areas where the local Euro-American dialect is *r*-pronouncing. Many American cities now harbor large Hispanic or Latino populations as well. Their speech has been less extensively studied than AAE (see Fought (2003) on the speech of Chicanos in California), but tends not surprisingly to feature varying degrees of substrate influence from Spanish. As they acquire English, upwardly mobile Latinos tend to converge with the sound qualities of SNAE, rather than with more

distinctive local dialects. Finally, both Canada and the United States are home to substantial Aboriginal or Indigenous populations, called Native Americans in the United States and First Nations peoples in Canada. Aboriginal English has been even less frequently studied than Latino English, despite its relative prominence in parts of the North American West and North, where the largest groups of Indigenous people live. Its phonology, however, tends to be fairly similar to that of SNAE, with only a few minor differences reflecting non-English substrates.

REFERENCES

Allen, H.B. 1959. Canadian-American speech differences along the middle border. *Journal of the Canadian Linguistic Association* 5(1): 17–24.

Avis, W.S. 1956. Speech differences along the Ontario–United States border. III. Pronunciation. *Journal of the Canadian Linguistic Association* 2(2): 41–59.

Baranowski, M. 2007. *Phonological Variation and Change in the Dialect of Charleston, South Carolina. Publication of the American Dialect Society No. 92*, Durham, NC: Duke University Press.

Bloomfield, L. 1933. *Language*, Chicago: University of Chicago Press.

Boberg, C. 2000. Geolinguistic diffusion and the U.S.–Canada border. *Language Variation and Change* 12: 1–24.

Boberg, C. 2001. The phonological status of Western New England. *American Speech* 76: 3–29.

Boberg, C. 2010. *The English Language in Canada: Status, History and Comparative Analysis,* Cambridge: Cambridge University Press.

Boberg, C. 2012. Ethnic dialects in North American English. In: *Rethinking Approaches to the History of English*, T. Nevalainen and E. Traugott (eds.), 538–548, Oxford: Oxford University Press.

Boberg, C. and S. Strassel. 2000. Short-a in Cincinnati: a change in progress. *Journal of English Linguistics* 28: 108–126.

Chambers, J.K. 1973. Canadian Raising. *Canadian Journal of Linguistics* 18(2): 113–135.

Clarke, S. 2004. Newfoundland English: phonology. In: *A Handbook of Varieties of English*, B. Kortmann and E.W. Schneider (eds.), vol. 1, 366–382, Berlin: Mouton de Gruyter.

Clarke, S. 2010. *Newfoundland and Labrador English,* Edinburgh: Edinburgh University Press.

Clarke, S., Elms, F., and Youssef, A. 1995. The third dialect of English: some Canadian evidence. *Language Variation and Change* 7: 209–228.

Dodsworth, R. and Kohn, M. 2012. Urban rejection of the vernacular: the SVS undone. *Language and Variation and Change* 24: 221–245.

Fought, C. 2003. *Chicano English in Context*, New York: Palgrave Macmillan.

Fridland, V. 2001. The social dimension of the Southern Vowel Shift: gender, age and class. *Journal of Sociolinguistics* 5: 233–253.

Frazer, T. 1978. South Midland pronunciation in the North Central states. *American Speech* 53: 40–48.

Gregg, R.J. 1957a. Notes on the pronunciation of Canadian English as spoken in Vancouver, B.C. *Journal of the Canadian Linguistic Association* 3(1): 20–26.

Gregg, R.J. 1957b. Neutralisation and fusion of vocalic phonemes in Canadian English as spoken in the Vancouver area. *Journal of the Canadian Linguistic Association* 3(2): 78–83.

Habick, T. 1993. Farmer City, Illinois: sound systems shifting south. In: *"Heartland" English*, T.C. Frazer (ed.), 97–124, Tuscaloosa, AL: University of Alabama Press.

Kennedy, R. and Grama, J. 2012. Chain shifting and centralization in California vowels: an acoustic analysis. *American Speech* 87: 39–56.

Kretzschmar Jr., W.A. 2004. Standard American English pronunciation. In: *A Handbook of Varieties of English*, B. Kortmann and E.W. Schneider (eds.), vol. 1, 257–269, Berlin: Mouton de Gruyter.

Johnstone, B. and Kiesling, S.F. 2008. Indexicality and experience: exploring the meanings of /aw/-monophthongization in Pittsburgh. *Journal of Sociolinguistics* 12: 5–33.

Joos, M. 1942. A phonological dilemma in Canadian English. *Language* 18: 141–144.

Kenyon, J.S. and Knott, T.A. 1953. *A Pronouncing Dictionary of American English*. Springfield, MA: Merriam-Webster, Inc.

Kurath, H. and McDavid, R.I. 1961. *The Pronunciation of English in the Atlantic States*. Tuscaloosa, AL: University of Alabama Press.

Labov, W. 1963. The social motivation of a sound change. *Word* 19: 273–309.

Labov, W. 1966. *The Social Stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.

Labov, W. 1972a. *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Labov, W. 1972b. *Language in the Inner City: Studies in the Black English Vernacular*, Philadelphia: University of Pennsylvania Press.

Labov, W. 1991. The three dialects of English. In: *New Ways of Analyzing Sound Change*, P. Eckert (ed.), 1–44, New York: Academic Press.

Labov, W., Ash, S., and Boberg, C. 2006. *The Atlas of North American English: Phonetics, Phonology and Sound Change,* Berlin: Mouton de Gruyter.

Labov, W., Yaeger, M., and Steiner, R. 1972. *A Quantitative Study of Sound Change in Progress*, Philadelphia: U.S. Regional Survey.

Ladefoged, P. 2006. *A Course in Phonetics*, 5th edition, Boston: Wadsworth.

Laferrière, M. 1979. Ethnicity in phonological variation and change. *Language* 55: 603–617.

Marckwardt, A.H. 1957. *Principal and Subsidiary Dialect Areas in the North-Central States, Publication of the American Dialect Society No. 27*, Tuscaloosa, AL: University of Alabama Press.

Martinet, A. 1955. *L'Économie des Changements Phonétiques*, Berne: Francke.

McCarthy, C. 2011. The Northern Cities Shift in Chicago. *Journal of English Linguistics* 39: 166–187.

Murray, T.E. 1986. *The Language of St. Louis, Missouri: Variation in the Gateway City*, Bern: Peter Lang.

Reed, C.E. 1952. The pronunciation of English in the State of Washington. *American Speech* 27: 186–189.

Thomas, E.R. 1997. A rural/metropolitan split in the speech of Texas Anglos. *Language Variation and Change* 9: 309–332.

Thomas, E.R. 2001. *An Acoustic Analysis of Vowel Variation in New World English, Publication of the American Dialect Society No. 85*, Durham, NC: Duke University Press.

Thomas, E.R. 2010. A longitudinal analysis of the durability of the Northern-Midland dialect boundary in Ohio. *American Speech* 85: 375–430.

Wells, J.C. 1982. *Accents of English*, Cambridge: Cambridge University Press.

Wolfram, W. 1969. *A Sociolinguistic Description of Detroit Negro Speech*, Washington, DC: Center for Applied Linguistics.

Wolfram, W. and Schilling-Estes, N. 1997. *Hoi Toide on the Outer Banks: The Story of the Ocracoke Brogue*, Chapel Hill, NC: University of North Carolina Press.

14 British English

CLIVE UPTON

The state of British English pronunciation

British English pronunciations range along a cline from the most regionally marked to that accent generally known as Received Pronunciation (RP), that is at least within England where it is essentially region-neutral. The multiplicity of accents in the British Isles stems in part from social structures, resulting in placement on the cline depending on speakers' needs and on the pressures exerted upon them to conform to some group norm. In larger measure, however, the variety of accents is born of the sheer length of time, some sixteen centuries, over which the language has developed in the islands. This "time depth" has led to greater fragmentation of speech-forms than that which has yet occurred in other places to which English has subsequently migrated: different influences have been exerted on the language and local allegiances have built up, to the extent that variety has become inevitable and is greatly cherished as a signal of regional and social identity.

This is not to say that RP is little regarded. Many British people both admire the accent and are pleased to see it having international status. However, comparatively few people speak RP, even that variety of it that is quite unmarked for privilege (see below on variation within the RP accent). Estimates vary on RP use in England. Wells (1982: 118) puts the figure at 10% "[e]ven with the more generous definitions [of what constitutes the accent]", while Romaine (2000: 20): put the figure at 5% at most. Such figures are purely guesstimates in fact, since no objective research into the matter has been carried out, and if all varieties of RP are counted together even Wells' figure might be rather low. Nevertheless, most people are readily identifiable to a place or region, which they would not be were they only RP speakers, and many people who have access to RP as one style of pronunciation do have access to more regionally-identifiable pronunciations too, which they use at need naturally and unconsciously to accommodate to more localized speaker situations. So the picture is complicated as regards the

The Handbook of English Pronunciation, First Edition. Edited by Marnie Reed and John M. Levis.

kinds of pronunciations that are to be heard in the British Isles. On the one hand there is a quite regionless, and now fairly classless, accent, RP. There is a multiplicity of regionally and still more locally espoused pronunciations, which are used by the majority of people all or most of the time. Most speakers roam, with greater or lesser ease, between accents at or approaching RP and accents that are very readily identifiable non-RP, which are sometimes regional to a very marked degree.

There is a long tradition of describing and analysing RP, beginning with Daniel Jones in the very early twentieth century (Jones 1909, 1917) and moving into the present day with a variety of materials, including many directed at the English-teaching and -learning community. RP models are also set out in pronouncing dictionaries of different styles and with varying perspectives on how the model is to be described (see, for example, Jones 2011; Olausson and Sangster 2006; Upton, Kretzschmar, and Konopka 2001; Wells 2008). Likewise, there exists a wealth of authoritative descriptions and analyses of regional and social British accent variations. Many of these are monographs dealing with the pronunciations to be found in particular regions of Britain. Stuart-Smith (2003) is an excellent example. Others are in the form of overviews of accent variety more generally. They might be written for the scholarly community. Wells (1982) is a model here. Alternatively they might be aimed at instructing the early student. Hughes, Trudgill, and Watt (2005) and Trudgill (1999) are good examples. Longer, more analytical and discursive treatment based on large research data are also available, some of this being available online. A most notable collection is to be found on the Accents and Dialects site of the British Library (http://sounds.bl.uk/Accents-and-dialects). At the historical level we might remark the many publications resulting from the Survey of English Dialects (Orton et al. 1962–1971), which are used to inform most commentaries, and the Linguistic Survey of Scotland (Mather and Speitel 1975–1986). These English and Scottish surveys have continuing relevance in the British sections of Schneider et al. (2004), where their findings emerge alongside more recent research to inform our understanding of regional pronunciation distributions. This work, being a compendium of very recent research by leading scholars of regional accents in the British Isles, is recommended as an ultimate authority. Its scholarship, also available in Kortmann and Upton (2008), furnishes in amplified form that which is digested here.[1]

A model English accent: Received Pronunciation

More than one Received Pronunciation

Received Pronunciation (RP) might seem a straightforward concept. It is to be found, usually without critical explanation, question, or qualification, as the exemplum in countless ELT books. It has most significantly been used for establishing the "standard lexical sets" system found in Wells (1982), the sets being "based on

the vowel correspondences which apply between British Received Pronunciation and (a variety of) General American" (Wells 1982: xviii). This system has become something of an "industry standard" for the discussion of English vowels, and it is used in this chapter. It would be comforting to think of RP, then, as a fixed point of reference for description and teaching. However, nothing relating to the accent is entirely straightforward. At an elementary level of description, we must first recognize that RP only relates to an accent of England: it is *English*, not *British*. This is important, as the other components of the British Isles, Wales, Scotland, and Ireland (described in more detail below), have, alongside their regional variations of pronunciation, variants that can to some considerable extent be considered as "standards", and which are widely regarded as such within the British Isles. It will be apparent to listeners to national radio and television that today even newsreaders, who might once have been expected formally to address their audiences in RP, are possessed of accents far removed from this, and that Welsh and Scottish accents especially are often to the fore. Of course, those accents will have been selected to be readily comprehensible to a wide international audience, but they will differ from RP to a marked degree in the regional elements that they contain. No one hearing an authoritative voice from Britain, therefore, should assume that they are hearing an RP accent. More significantly still, no one hearing an *English* voice should assume this either. There is today greater acceptance of the regional accents of England in broadcasting and the professions than there was formerly, rendering the identification of RP uncertain and, it must be said, of rather questionable importance, for native British English speakers themselves.

Even if one targets RP as a desirable acquisition, it will be apparent to the linguistically aware that the notion of it being one immutable accent must be a fiction. This iconic variety is a moving target on which its describers and teachers have constantly to adjust their aim. But in spite of there being a variety of labels referring to the same and different subvarieties (Upton 2004: 219–220), not to mention re-description reaching to the *Oxford English Dictionary* itself (Upton 2012: 62–65), experience shows that many people, including some language professionals, have a fixed notion of how the RP model is to be transcribed, and that it is firmly rooted in the past. A complication is the fact that, faced with linguistic change, people will accommodate to it at different speeds and to different degrees. Communities will therefore have a range of speakers possessed of accents ranging from the progressive to the reactionary, with a ready sprinkling of those whose idiolects show signs of misunderstanding of or indifference to what actually qualifies as any kind of RP. The result is that transcriptions of what is claimed to be RP, if faithfully reproduced according to phonetic principles, can lead to pronunciations that will sound old-fashioned, quaint, or even affected to many native British English speakers. By the same token, attempts to revise formal descriptions of RP will be met with incomprehension by some and, pronunciation being a contentious subject, with hostility or even outrage by more than a few, who either misunderstand what is being attempted or are simply resistant to the notion that an iconic model will change. (It is ironic that those who are resistant to the notion that RP can change are themselves the inheritors of a description model that is in some important

respects markedly different from Jones' "Public School Pronunciation" (PSP) and the RP that it quickly became. (See Upton 2012: 58-60 for details of early variants later superseded.)

Today's RP, a re-description not a revolution

Happily, the variables that are contentious in a (re)description of RP today, though individually significant, are few in number. Most of the variants of the model accent that, following Ramsaran's (1990) labelling "traditional RP", I have elsewhere (Upton 2004) termed "trad-RP", continue into the present. There is therefore considerable coincidence between the transcriptions of Wells (2008) and Jones (2011), on the one hand, and those of Upton (2004 [and the ongoing Oxford Dictionaries]) and Olausson and Sangster (2006), on the other, which seek some modest re-design.

RP vowels

Table 14.1 charts vowel transcriptions that are most generally encountered in available descriptions of RP today. The vowel column makes use of the keywords of the Wells (1982) system of lexical sets. The RP column conveys the vowel transcriptions available in the Br[itish] element of Upton, Kretzschmar, and Konopka (2001) and in Olausson and Sangster (2006), and (alongside North American equivalents) in the online *OED* third edition. The trad-RP column shows those points at which more traditionally conservative systems of RP vowel transcription differ from those of RP proposed here, and the notes briefly explain those differences. The discussion that follows concentrates only on areas of difference. Areas that show no difference are not commented upon.

DRESS: When first described, this RP vowel was at or near half-close, rendering [e] the natural phonetic choice. The vowel has now tended to be lowered to a point at or near half-open, rendering [ɛ] the more accurate choice of symbol. The [e] transcription is apparently favoured by some transcribers for reasons of continuity.

TRAP: When studying English pronunciation, look and listen for [æ]. As with DRESS, the RP vowel has lowered, a point made as long ago as 1982 by Wells (1982: 292). Where [æ] is retained to describe the modern sound, this can now only be as a convention, and Cruttenden (2014) has the [a] transcription.

BATH: No one questions the use of [ɑː] for BATH. It is clear, however, that very many speakers in Midland and Northern England accord with their southern neighbours in all RP pronunciations except this, where [a] is very frequently used by them instead. The contention here is that RP should not be regarded as a southern English accent merely on the basis of one sound, with very many speakers consigned to a "near-RP" category because they diverge systematically from an established pattern by one distinction. It is therefore reasonable to recognize two variants for this variable. I have written a justification for this "mould-breaking" innovation in transcription elsewhere (Upton 2012: 64–65).

Table 14.1 Modern unmarked vowel transcription for Received Pronunciation, with present-day transcriptions of traditional alternatives. Adapted from Upton (2004).

Keyword	*RP*	*or*	*trad-RP*	*Note*
KIT	ɪ			
DRESS	ɛ		e	trad-RP symbol kept conventionally
TRAP	a		æ	trad-RP symbol kept conventionally
LOT	ɒ			
STRUT	ʌ			
FOOT	ʊ			
BATH	ɑː~a		ɑː	Short vowel in northern RP
CLOTH	ɒ		ɒ~ ɔː	Long vowel only in the most rarified trad RP
NURSE	əː		ɜː	Symbol difference only
FLEECE	iː			
FACE	eɪ			
PALM	ɑː			
THOUGHT	ɔː			
GOAT	əʊ		əʊ~oʊ	trad-RP [oʊ] variant might be resurgent
GOOSE	uː			
PRICE	ʌɪ		aɪ	Difference largely symbolic
CHOICE	ɔɪ			
MOUTH	aʊ			
NEAR	ɪə			
SQUARE	ɛː		ɛə~eə	Some off-gliding, rarely full diphthong
START	ɑː			
NORTH	ɔː			
FORCE	ɔː			
CURE	ʊə~ɔː		ʊə	RP monophthong increasing
happY	i		ɪ	[ɪ] very conservative only
lettER	ə			
commA	ə			

CLOTH: The [ɔː] variant is very recessive, and seems risible, or at very best old-fashioned, to the majority of native British English speakers.

NURSE: It is apparent from the evidence of those using the [ɜː] transcription that the mid-central position is indicated for NURSE and that this symbol is simply conventionally deployed to disambiguate it from the unstressed commA vowel. However, use of [əː] creates no ambiguity and reduces the symbol set needed for RP description by one.

GOAT: It is today generally agreed that the onset of RP GOAT is best regarded as being [ə]. It has, however, formerly been retracted, giving [oʊ]. There is a strong

possibility that the GOAT vowel may return to a less centralized onset in the near future.

PRICE: This is undoubtedly the most controversial of the RP re-descriptions attempted in recent years. Problematic is the use of [a] for onset, since if it is taken to imply low-front articulation it results in pronunciation of a very traditional diphthong. Low-front articulation is not what is implied by use of this symbol in much conventional transcription, however, as it now tends to suggest a retracted or centralized onset. At the same time, RP /ʌ/ is generally agreed to be centralized and below half-open, justifying selection of this as the onset vowel for PRICE. The lexical trio <fan~fun~fine> is instructive here: those attempting RP <fine> would do well to move from <fun> ([fʌn]) to <fine> ([fʌɪn]) rather than going from <fan> ([fan]) to <fine> ([faɪn]), assuming of course that they use today's RP realization of /ʌ/.

SQUARE: Conventionally transcribed as diphthongal, RP SQUARE is essentially in fact a monophthong. A recent adoption of this is in Cruttenden (2014). Some slight off-gliding might be identified, but full diphthongization here results in a sound that seems dated to most British native-speaker audiences.

CURE: Long-established and increasingly heard today is monophthongal [ɔː] in place of [ʊə], especially though not only amongst younger speakers. It is likely that [ʊə] will be little heard and that [ɔː] will be the norm for RP in future.

happY: [i] here, replacing former [ɪ], implies both tenseness and a degree of length, though not full length. The short lax vowel is only heard amongst a small set of generally older speakers and is strongly recessive as an RP feature.

RP consonants

As above for vowels, concentration here is on those issues of RP consonantal articulation that diverge from widely held notions.

In relaxed, informal speech yod coalescence is to be expected. Hence /sj/ can go to /ʃ/ in *assume*, /zj/ to /ʒ/ in *resume,* /tj/ to /tʃ/ in *Tuesday*, and /dj/ to /dʒ/ in *due*. As the second pair of examples here makes clear, coalescence might be expected word-initially as elsewhere in some words. Yod deletion has long been usual in words such /sj-/ words as *suit* (/suːt/) and, although not yet frequently heard, it is beginning to pass without remark in such a word as *news* [nuːz]).

Although RP is nonrhotic, both "linking *r*" (*here and there* /hɪər n ðɛː/) and also "intrusive *r*" (*drawing* [ˈdrɔːrɪŋ]) are normal, although their avoidance is a notable feature of trad-RP. As will be apparent from the *here and there* example, syllabic consonants are often to be encountered in RP, including for the conjunction *and*. Jones (1969: para. 213) sees syllabification as particularly a function of the "more sonorous consonants such as **n**, **l**".

Whilst glottalization is not an especially marked feature of RP, /t/-glottaling especially is by no means as unheard in the accent as is sometimes thought. It might particularly be expected syllable-finally preceding a nonsyllabic consonant, [rʌɪʔˈwɪŋə] *right-winger*. It will sometimes be found between vowels at a syllable boundary, where the first syllable is unstressed and the second stressed: [rɪˈʔɔːgənʌɪz] *reorganize*.

Locating regional accents

Whole accents do not map on to regions

RP is taken as a yardstick in the description of accents that follows, for no other reason than that this permits the omission of repetition should non-RP accents coincide with RP in certain particulars, although it must be appreciated that coincidence of a feature in RP and a localized accent *does not* mean that its user is speaking with an RP accent. Concentration from here on is primarily on *non*-RP sounds that have connections with particular parts of the British Isles, RP itself being drawn into the descriptions where inclusion might be informative. It is by such sounds as those that are *individually* associated with particular areas of the British Isles that speakers can be placed geographically. Particularly when they coincide with other sounds that are similarly located, they enable an informed hearer to identify a person's origin, or at least the principal influences that have acted upon their accent. In a situation such as that existing in the British Isles, where varieties abound and many speakers are socially and geographically mobile, accents do not, of course, occur in tidy, monolithic blocks, each block distinct from another. Rather, a community, and indeed each speaker within a community, will exhibit features drawn from a wide area in the creation of their unique accent. Each of the phonemes (or variables) of a language has a particular distribution pattern for its variants across a territory: each sound will loosely occupy its own geographical space, the distribution patterns for no two variables coinciding absolutely.

So, since it is not possible to isolate an entire set of sounds and to allocate them to a particular place or region, concentration here is on the attachment of individual features to regions, these features being discussed one by one, rather than making an attempt to identify an amalgam of specific features all coming together in one place. Since it is the vowel system that is most telling of place, the primary device used in analysis is again that of standard lexical sets. Reference should be made to the country-wide (and RP) realizations from which divergence is described here. The descriptions are necessarily truncated, although they do provide what is needed for the reader to begin to form a proper understanding of sounds that a native British speaker is likely to use when seeking to identify a speaker with a place. The major resources drawn upon for the selection of these features are Schneider et al. (2004) and Kortmann and Upton (2008), and the reader is recommended to move out from here to those works in order to flesh out the thumbnail descriptions.

The "British Isles" and their parts

In order that a broad overview of the geographical distribution of major phonological distinctions might be followed, it is necessary to provide some short explanation of terms used to relate to areas of the British Isles. To begin with what is for many a particularly problematic geopolitical issue, the very term "British Isles"

has to be addressed. It refers to all those islands that contain two adjacent but quite distinct nation states, the Republic of Ireland (or Eire) and the United Kingdom of Great Britain and Northern Ireland. The Irish Republic occupies the south, middle, and north-west of the large island of Ireland, on the western side of the region. The United Kingdom (UK) takes in the countries of Scotland, Wales, Northern Ireland (or Ulster) occupying the north-eastern part of the island of Ireland, and England.

In what follows, the major designations that will be encountered will be "Ireland", by which is meant the whole of the island that bears the name, "Scotland" in the north of the UK, "Wales" in the west, and "England". Upon occasion it is necessary to distinguish between "Southern Ireland" and "Northern Ireland" with an implication that forms relate essentially to the Republic (south) or to the Irish part of the UK, Ulster (north). Wales and Scotland are also referred to separately, upon occasion with compass-based geographical subdivisions. Archipelagos extending northwards from the Scottish mainland, which exhibit markedly distinct characteristics for some variables, are the "Orkney Islands" and the "Shetland Islands" (together the "Northern Isles"). Descriptions within England most essentially see the country divided into "north", "south", and "Midlands", this last separating again into the "East Midlands" and the "West Midlands". The Midlands constitutes a transitional zone of indeterminate breadth exhibiting both shared northern or southern and region-specific features. The most easterly part of the East Midlands, which exhibits very distinctly-heard variants for some variables, is identified as "East Anglia". Within the North we must at times identify as distinct the "North-east", an area centered on the city of Newcastle upon Tyne and abutting the Scottish border. The south of England is upon occasion separated into the "south-west" (sometimes referred to as the "West Country") and the "south-east", an area dominated linguistically to some considerable extent by London. To the south of the English mainland the "Channel Islands", with their historical French influence, again warrant separate mention at times.

Major markers of place

/r/

The most significant pointer to broad regionality applying within the British Isles is that of rhoticity, the pronouncing of /r/ following a vowel, where <r> occurs in a word when written. (Rhoticity is signalled in English spellings that were fixed before many English accents became nonrhotic, this happening comparatively recently, with local speech even close to London evidencing the feature until the middle of the twentieth century (Orton et al. 1962–1971).) Rhoticity is a worldwide phenomenon, being, for example, a feature that predominates in the English pronunciations of North America, and is more common than is often supposed within the British Isles. It is the norm in Ireland, Scotland (though apparently receding in some urban areas, notably Glasgow), and in parts of Wales (in the south-west of

the country and by transference from their Welsh in the English of Welsh speakers). Within England it is found in the West Country, southern Lancashire, and (as essentially a feature of older people's speech) in the far north-east, north of Newcastle upon Tyne. Rhoticity has an effect on many preceding sounds: commonly in Scottish English there is no diphthong in such as word as *here* and *sure*, these being pronounced [hir], [ʃur].

Whether associated with rhoticity or not, realizations of /r/ are variable throughout the British Isles. Within Ireland an essential difference is between southern [ɹ] and northern [ɻ], though with a spreading of an unrelated [ɻ] outwards from the capital city Dublin in the south as well. The reverse of this is true of England, where [ɻ] is south-western and [ɹ] is more usual elswehere. The north-east England occasional feature is the "Northumberland Burr", [ʁ]. In Wales, [ɾ] is especially brought from Welsh by Welsh-speaking speakers when they are using English. Scottish /r/ ranges from post-alveolar to retroflex or a tap.

/əː/ NURSE

Further to an absence of centring diphthongization in Scotland (as described under rhoticity), many Scottish English speakers do not have [ə] in their inventories, substituting another vowel instead at need. As a result we might expect *nurse* itself to be rendered [nɛrs] or [nʌrs]; in the absence of [ə], the English hesitancy form [əː] is quite typically [eː] in Scotland. Indicative of (especially south) Welsh English is [œː]~[øː], giving [bœːd~bøːd] *bird*, [tœːn~tøːn] *turn*. There is merging of NURSE with NORTH on [ɔː] in the speech of some older, especially rhotic, speakers in north-east England, but [øː] is to be heard from some of the youngest speakers.

/a/~/ɑː/ BATH

Variation in the BATH vowel before following /s/, /f/, or /θ/ is, with STRUT /ʊ/~/ʌ/ variation (below), one of two features that are especially prominent characteristics of a north–south division within England. As indicated in the RP discussion for this feature, the principal variability in the vernacular is between [a] in the north and [ɑː] in the south, with the Midlands split: the isogloss separating the two variants runs roughly due eastwards from the Wash (the large indentation on the English east coast north of East Anglia), through Birmingham, to the Welsh border. The [a] region in fact extends to this line from the far north of Scotland. As mentioned, so firmly fixed is this distinction that even the dictates of RP do not strongly challenge regional allegiance for northern RP speakers. Historically, earlier country-wide BATH [a] first became lengthened in the south to [aː], and this is still heard outside the south-east. The north–south distinction in England is thus essentially one of length, north/short and south/long, with vowel quality in southern regions a lesser issue: essentially [aː] is indicative of the south-west and East Anglia, [ɑː] of the south-east. Southern England's (and "southern" RP"s) [ɑː] is a later development of this; [aː] is characteristic of Wales and of Southern Ireland, where it contrasts with Northern Irish [ɑ].

/ʊ/~/ʌ/ STRUT

The principal distinction for STRUT within the dialects of England specifically is between /ʌ/ (as in RP), found in southernmost areas of the country, and the more historically grounded /ʊ/ of the North and Midlands. This, like the BATH distinction above, is used by very many people to identify speakers as either northern or southern English. However, the isogloss separating the variants does not coincide with that for the BATH variation running straight across the Midlands. Rather, it dips southwards at its mid-point to reach the Thames Valley west of London: given that the BATH isogloss takes a more northerly track, running straight west-to-east across the middle of the country, it would be quite wrong to think that there is a clear "north–south divide" in pronunciation in England founded on BATH and STRUT (or any other variables). Also, unlike the /a/-/ɑː/ north–south distinction referred to in the discussion of RP, the use of STRUT /ʊ/ is the subject of more critical public remark, perhaps because, rather than it being a mere phonetic difference, it signals the absence of an entire phoneme, /ʌ/, from a northern speaker's phonemic inventory. Most likely as a consequence of this there exists an intermediate STRUT sound, [ɤ], dubbed a "fudge" by Chambers and Trudgill (1998: 110; see also Upton 1995: 385–394) since it occurs between the two alternatives in articulation, and hence acoustically. This fudge is found widely, especially though not exclusively at the interface between areas of strong [ʊ] allegiance (the North) and those favouring [ʌ] (the south-east especially). [ʌ] tends to be centralized particularly in Wales and Ireland, while [ɔ] is a feature of the accents of the Channel Islands and Orkney and Shetland.

/eː/~/eɪ/ FACE

Dubbed "long mid diphthonging" by Wells (1982: 210–211), the fracture of the long monophthong to a diphthong here at FACE, and also at GOAT (below), is a historical process that has been variously applied in the British Isles. Along with GOAT among diphthongal vowels, and especially BATH and STRUT among monophthongs, FACE is particularly distinguished by a north–south split, in this case (and with GOAT) involving a monophthongal north from the North of England northwards through Scotland, and in Wales and Ireland, contrasting with a diphthongal Midlands and southern England. So [eː] is typical in the North of England, Scotland, Ireland, and Wales; /eɪ/ is to be expected from the English Midlands southwards, with [ɛɪ]~[ʌɪ]~[æɪ] realizations also found there. One notable exception to a largely monophthongal North is historical [ɪə] in north-east England, now as a somewhat recessive feature.

/oː/~/oʊ/ GOAT

The distribution of long monopthong versus diphthong is found here as is found for the FACE vowel. Accents of the South and Midlands of England, in the west extending as far north as Liverpool, are characterized generally by diphthongs, the most significant ones other than frequent [oʊ] being [ʌʊ] in the south-east, [ɔʊ] in the south-west, and [ʌʊ~aʊ] in the Midlands. In East Anglia there is variability in GOAT, between [ʌʊ] and [ʊu], according to etymology. See Wells (1982: 337) for an explanation of this.

In contrast to Southern and Midland English diphthonging, the monophthong [oː] is quite characteristic of the accents of Northern England, Scotland, Ireland, and Wales. Basic monophthong-diphthong GOAT variation, like that for FACE, is thus used by many listeners to place a speaker geographically. As noted in the section on RP above, trad-RP [oʊ] seems to be resurgent, a fact that might well be linked to its presence in some regional accents in England and Wales, and to some extent in accents outside these regions.

A social development in the North, spreading especially as a feature of the speech of younger middle class from the north-east and the east coast city of Hull is "GOAT-fronting" to [øː]. A recessive feature of note is the North-east England traditional [ʊə], paralleling [ɪə] in FACE.

/h-/

It is often thought that "h-dropping" or "h-deletion", resulting in [aʊs] *house*, [apn] *happen*, is a universal feature of non-RP in Britain, but this is not the case. Whilst it is widespread in Wales, it is unusual in Scotland and Ireland, and although it is frequent in large areas of England it is not usual in the rural areas of East Anglia, or in the north-east north of Newcastle. Tied to matters of orthography as it is, the dropping of the initial [h-] tends to be socially stigmatized, rendering it the subject of as much sociolinguistic as it is regional-distributional debate (see, for example, Mugglestone 2003: 95ff).

A longstanding spelling signaling a pronunciation that is now regionally associated is <wh->, representing [ʍ-]. Whilst this is found in somewhat mannered forms of RP, it is especially associated with pronunciations local to Ireland, Scotland, and the Scottish–English borderland. In Ireland the pronunciation chimes well with a similar sound in the Irish language.

/aʊ/ MOUTH

A pronunciation especially to be expected in Scotland, but also extending southwards into north-east England, is [uː] for MOUTH. As is the case with many other regionally significant features, this feature is of very considerable antiquity, and, like PRICE [iː], was the norm for English before the onset of the Great Vowel Shift of the late Middle Ages (Smith 1996: 86-111). Although it is not as ubiquitous as it once was in general use, especially to the south of the Scottish border, it retains special currency in north-east England as a marker of local identity. Newcastle United football club, for example, is frequently referred to as *The Toon* (i.e., "The Town", [ðə tuːn]), this being used as a signal of local allegiance by many people who would nevertheless regularly use MOUTH [aʊ].

Forms of /l/

Helpful distinctiveness here is between "clear" or "thin" [l] and "dark" or "thick" [ɬ]. There is a trend from clear to dark as one moves south through England, with full vocalization to [ʊ] being found in the south-east around London. Quite the opposite trend occurs in Wales, with the thin variant in the south, the thick variant being found in the north, as it is in Scotland.

Fine tuning regional differences

/ʌɪ/ PRICE

The RP diphthong [ʌɪ] is shared with local accents widely in Scotland, while the trad-RP [aɪ] is also found in Orkney and Shetland, Northern (often with a lengthened onset) and Midland England, Wales, and Southern Ireland. Higher onsets for the diphthong, giving [æɪ]~[ɛɪ], can be characteristic of rural Irish accents. Low-back onsets are also widely heard, as [ɑɪ] in Southern England and the Channel Islands, East Anglia, and in Ireland, especially in Dublin, and as [ɒɪ] in the West Country and West Midlands of England, and London (Cockney).

Originally the norm in the PRICE set was [iː]. This, like MOUTH [uː] above, a form dating from before the time of the medieval Great Vowel Shift, has become lexicalized, especially in the Yorkshire region of northern England. Here especially it operates restrictedly but significantly in a small set of words, most notably *right* and *night*, to signal local affiliation.

Representations of <-ng>

A feature now exhibiting quite restricted distribution is "velar nasal plus", the insertion of the alveolar stop [g] following [ŋ]. This results in *long* being pronounced [lɒŋg], *thing* [Θɪŋg]; *singing*, which is in many accents [ˈsɪŋɪŋ], becomes with this feature [ˈsɪŋgɪŋ] or even [ˈsɪŋgɪŋg]. Formerly widespread amongst English speakers, velar nasal plus is now a quite reliable indicator that a speaker comes from the north-west Midland region of England, and to be located somewhere between Birmingham and southern Lancashire/Yorkshire.

Like absence of [h-], [-n] in words with <-ing>, such as *coming*, is a socially stigmatized feature of pronunciation, and so is the subject of sociolinguistic study. It does not manifest geographical distribution, however, being found widely across the whole of the British Isles.

Distinctive treatment of stops

Like velar nasal plus a feature of the north-west Midlands, but restricted to its extreme northern edge, is Liverpool affrication of the voiceless stops /p/, /t/, and /k/, these becoming respectively [ts], [pf], and [kχ], and in final position sometimes realized as fricatives [ɸ, s, χ]. Seemingly as a consequence of this affrication, glottaling of /t/ to [ʔ], which is to be found in almost all British accents, is less usual in Liverpool than elsewhere.

Unlike the widespread /t/-glottaling, however, glottaling of /p/ and /k/ is particularly a feature of north-east England, this being increasingly heard from younger speakers to replace the glottalization (pre-glottaling) forms [ʔt, ʔp, ʔk], which were formerly ubiquitous and which are now typically heard from more elderly northern England speakers.

/ɒ/ LOT

The major difference from an RP-like variant is [ɔ] in Scotland (including the Northern Isles), in Wales, and in the West Midlands. It is also a feature of

"fashionable" speech in Dublin, where it contrasts with [a] in more colloquial speech found there and in rural areas in the west of the country; [ɑ] is also heard in Southern Ireland, and in south-west England and East Anglia.

/ɒ/ CLOTH
The lengthened form [ɔː] found in trad-RP is also a feature in the accent of older speakers in East Anglia. As with LOT, [ɑ] can be heard especially in Southern Ireland, the West Country, and East Anglia.

/ʊ/ FOOT
Accents in Scotland and Ireland tend towards a tense close central rounded [ʉ] here, sometimes advanced to a close front rounded [Y]. [Y] is also a feature of the English West Country. Long-standing stigmatization of Northern English STRUT [ʊ], especially imposed on this area from without, can result in realization of FOOT as [ʌ] through hypercorrection.

/uː/ GOOSE
[uː] is ubiquitous here in regional accents as in RP, with an advanced form [ʉː] frequently heard too (increasingly in emerging forms of RP as regionally). An on-glide on [u] is especially frequent in the West Midlands, giving [əu(ː)], with shorter diphthongs heard in the north of England, East Anglia, and south-east England, [ʊu]. Especially older speakers in the West Country might exhibit [Yː], with [Y] heard in urban Scots.

/iː/ FLEECE
Rather than this being universally a long monophthong, there is a tendency for there to be a short diphthong, [ii], here, with wider diphthongs based on [əɪ] found especially in England. As with GOOSE, on-gliding is particularly notable in the West Midlands, giving [əiː].

/a/ TRAP
As in RP, [æ] is also being replaced by [a] in regional southern English accents, though it is still regularly heard there, as it is in Ireland, East Anglia and the East Midlands, and the Channel Islands; [a] is the norm elsewhere, often in a retracted form in urban Scotland.

/ɑː/ PALM, START
RP-like [ɑː] is found in south-east England and the Channel Islands, and in northern England alongside [ɒː] (also in the West Midlands) and [aː]. This [aː], incidentally the immediate ancestor of RP [ɑː], is the norm in south-west England and in Wales.

Most variation relating to START coincides with that for PALM. However, while PALM tends to be short [a] in Scotland, there exists a rule, the "Scottish Vowel Length Rule", which states that Scottish vowels are long before fricatives, /r/, or at boundaries: this results in the Scottish [a] being lengthened to [aː] in the rhotic environment of START, which is sometimes retracted from low front or raised to [ɛː].

/ɔː/ NORTH, FORCE, THOUGHT
The most striking regional feature for NORTH and FORCE is the [ʊə] of north-east England, a historical feature that is declining in use but is characteristic of the most localized speech, speech that is sometimes rhoticized on [ʁ].

Otherwise distinctive for NORTH are [oː] in Scotland and this or [ɒː] in Northern England. In Ireland the range [ɔː~ɒː~ɑː] is found for NORTH. FORCE distinctions are to be heard in Ireland, where [oː~ɔː] are usual, with [ɒː] a Dublin feature.

THOUGHT exhibits a wide range of realizations. Principal exceptions to [ɔː] are the [ɔ] found in Scotland, [ɒː~aː] in northern England, and [oː~oʊ~ɔə] in the south-east of England. Irish speakers have a range of mainly back vowels, [ɔː~ɒː~ɑː], although [aː] is found in Dublin.

/ɪə/ NEAR
Nonrhotic accents of Britain, essentially those of England outside the south-west, mainly exhibit centring diphthongs with a high front onset at [ɪ] like RP, or tense [i]. West Midland accents might have [eə~ɛə]. Rhoticity brings with it a monophthongal realization on [i]. Monopthongal [eː]~[ɛː] can be found in nonrhotic East Anglia, creating a NEAR/SQUARE merger.

/ɛː/ SQUARE
The [ɛː~ɛə] of RP/trad-RP is also usual in nonrhotic accents (i.e., in most of England), though [ɜː] is characteristic of an area of the northern edge of the West Midlands centered upon Liverpool. Rhotic areas tend to have their rhoticity based on [e(ː)] or [ɛ(ː)], the latter typical of the West Country especially.

/ʊə/ CURE
The monophthongization that gives [ɔː] is increasingly the norm in RP and is found in Northern and Midland English regional accents also, although [ʊə] is most frequent in all the nonrhotic areas. The rhoticity characteristic of Scotland, Ireland, and south-west England is typically on [u(ː)]. Very telling of a Welsh English accent is a tendency to a disyllabic [uwə] or [ɪuwə].

/ɔɪ/ CHOICE
[ɔɪ] can be expected in all regions of the British Isles, though the glide tends to be somewhat lower in Scotland, at [ɔe]. Other low-back onsets producing [ɒɪ]~[ɑɪ] can be considered somewhat characteristic of the Irish Republic, the English North, and the West Midlands, where [oɪ] is also found; this [oɪ] also occurs in Dublin and south-east England. Raised onsets can also be heard in East Anglia, with [ʊɪ].

/ɪ/ KIT
Beyond [ɪ], which occurs in all varieties and is usual except in the islands to the north of Scotland (Orkney and Shetland), where there is a distinct tendency towards retraction and lowering, the main characteristic marker of regionality is tense [i]. This is a particular feature of the West Midlands, and is also to be heard in East Anglia.

/ɛ/ DRESS
Raising from the typical RP vowel to [e] is especially a feature of south-east England, and is also found in East Anglia, the Channel Islands, Orkney and Shetland, and the cities of Scotland. Lowering to [æ] is very readily to be heard in Northern Ireland.

Unstressed vowels
/i/ happY: While [i], generally with some element of lengthening, is particularly widespread, [ɪ~e~ɛ] are traditionally found in the North of England, [ɪ~e] in Ireland, and [e] in Scotland.

lettER: [ə] is generally found irrespective of the presence or absence of rhoticity. Rhotic Scotland also has [ɪ~ʌ]. Nonrhotic Wales can have [ʌ], and the Channel Islands [œ].

horsES: [ə], alongside [ɪ], is characteristic of northern England, and is usual in East Anglia and Ireland.

commA: Alternatives to the generally widespread [ə] here tend to involve lowering to [ɐ] in Shetland, Ireland (notably Dublin) and Northern England, and [ʌ] in Scotland and Wales.

Northern English speakers exhibit a tendency to give full value to vowels in unaccented syllables, so that, for example, *condition* might be rendered as [kɒnˈdɪʃn].

/Θ/ and /ð/
Very distinctive of Southern Ireland is the use of alveolar stops where most British English accents have the equivalent stops: hence /t/ for /Θ/ and /d/ for /ð/.

Some regional suprasegmentals

The present shortage of phonological research data beyond the segmental renders the making of detailed regional distinctions impossible. It is, however, possible to make some general observations in this area, using data from the contributors to Kortmann and Upton (2008).

Scottish intonation study has observed a high-rising pattern for statements and questions in Glasgow and a falling one in Edinburgh and elsewhere. The Glasgow phenomenon might be influenced by Northern Irish speech. Distinct from a high-rising pattern is a level high-rise intonation terminal that has been detected in north-east England. The South Wales valleys have attracted attention because of notable variation in pitch movement, with possible influence from Welsh, and this has been likened to a similar, though of course unconnected, feature in Orkney (though not apparently Shetland) English. East Anglia is notable for intonational movement from low to high levels during the asking of yes–no questions especially.

Salient in north-east England is a tendency to level stress or heavier second-element stress in compounds, so that the city of Newcastle (upon Tyne) is pronounced by many inhabitants [njʊˈkasl] (as opposed to [ˈnjuːkasl]). A similar feature of regular stressing in Channel Islands English seems likely to be explained by historical and continuing French influence. There is a tendency towards stress-shifting to long final vowels in polysyllabic verbs in Irish English, as in *testify* [tɛstɪˈfaɪ]. The lengthening of stressed vowels and the loss rather than reduction of unstressed vowels is a feature of East Anglian prosody, lending the variety a distinctive rhythm: *have you got a light?* [hæːjə gɑʔ lɐɪʔ]. Disyllabic words in Scotland show some tendency to have a short-long pattern of rhythm.

Conclusion

It has been asserted that variability is the inevitable state of accents. We cannot identify a set of sounds that can be easily allocated to one model of accent or to one particular territory. In consequence, even such a seemingly obvious institution as RP warrants discussion as regards variation. In its BATH-distinction discussed above, this supposedly regionless variety exhibits at least a small amount of location-based distinction. There also exists a range of tolerances between the most conservative trad-RP and the most progressive (or speculative) features, among which we should identify especially variants in the TRAP, PRICE, and SQUARE vowels. No one form of this "model" accent is "better" or "correct". Rather, its varieties are a decided sign of the vibrancy of the language of which it is representative.

Further to this theme of striking variability, it is impossible to identify a distinctive set of pronunciations that, together, ties a population uniquely to a place. It is the contention here that, in a very complex situation such as that which has arisen in the British Isles, where accents have had opportunity and cause to fragment over time, different variables have evolved uniquely in terms of their geographical distribution. It is indeed quite often possible for a listener to place a speaker as regards their region of origin, and also to detect in their speech other regional features to which they have been exposed and by which they have been influenced. However, it will be *particular* variants, used to realize just a few salient variables, that the listener will most often rely on when forming their judgments. A speaker will be placed as coming from Scotland, Ireland, or certain parts of England (the West Country, the lower north-west, the north-east) by the presence in their accent of post-vocalic /r/. The *type* of /r/ – fricative, retroflex, uvular – will determine matters more narrowly on this one feature alone.

It is not on a single feature that people are located, of course, but on an amalgam of those around which speakers cluster. Nevertheless, the more features that are aggregated the looser the community bonds become. A speaker from the central part of the English Midlands around Birmingham is likely to share "velar nasal plus" with a speaker from its Liverpool-centred northern edge. However, they are most unlikely to display any tendency to that stop-affrication that is so characteristic

of Liverpudlian speech: rather, they will in some considerable numbers at least use BATH [ɑː], tying them firmly to their neighbors further south. Most significantly, there is no identifiable point between Birmingham and Liverpool where the affrication feature is or is not solidly entrenched. Rather, its espousal as an accent feature will vary according to a complex of historical and present-day social factors, which affect each of the speakers across the north-west Midlands differently.

We have, then, in this one small example from within England, an important lesson, that accents, and indeed dialects, blur into one another.[2] We can place people roughly by their use of certain major accent identifiers. We can then spot pointers to smaller regions. But even speakers firmly rooted in one spot will not share all variants in the same proportions, and in societies that are increasingly mobile the combination of possible variants to which people have access multiplies. The detective work of narrowing down a speaker's origin in the British Isles is enthralling – just as long as one does not get frustrated when this proves elusive.

NOTES

1 I am grateful to Mouton de Gruyter for their express permission to make use of information from Schneider et al. (2004) and Kortmann and Upton (2008) throughout this chapter. The several scholars contributing to these works are the ultimate authorities to be consulted and I am entirely indebted to them for their invaluable information. They are: Ulrike Altendorf, Joan Beal, Urszula Clark, Raymond Hickey, Gunnel Melchers, Peter L. Patrick, Robert Penhallurick, Heinrich Ramisch, Jane Stuart-Smith, Peter Trudgill, and Dominic Watt.

2 See Davis and Houck (1995) and Davis, Houck, and Upton (1997) for exploration of the observation by Gaston Paris that "there really are no dialects".

REFERENCES

Chambers, J.K. and Peter Trudgill. 1998. *Dialectology*, Cambridge: Cambridge University Press.

Cruttenden, Alan. 2014. *Gimson's Pronunciation of English*, 8th edition, London: Routledge.

Davis, L.M. and Houck, C.L. 1995. What determines a dialect area? Evidence from the Linguistic Atlas of the Upper Midwest. *American Speech* 70(4): 371–386.

Davis, L.M., Houck, C.L., and Upton, C. 1997. The question of dialect boundaries: the SED and the American atlases. In: *Issues and Methods in Dialectology*, A.R. Thomas (ed.), 271–283, Bangor: University of Wales Bangor.

Hughes, A., Trudgill, P., and Watt, D. 2005. *English Accents and Dialects: An Introduction to Social and Regional Varieties of English in the British Isles*, 4th edition, London: Hodder Arnold.

Jones. D. 1909. *The Pronunciation of English*, Cambridge: Cambridge University Press.

Jones, D.. 1917. *An English Pronouncing Dictionary*, London: Dent.

Jones, D. 1969. *An Outline of English Phonetics*, Cambridge: Heffer.
Jones, D. 2011. *Cambridge English Pronouncing Dictionary*, 18th edition, P. Roach, J. Setter, and J. Esling (eds.), Cambridge: Cambridge University Press.
Kortmann, B. and Upton, C. (eds.). 2008. *Varieties of English 1: The British Isles*, Berlin: Mouton de Gruyter.
Mather, J.Y. and Speitel, H.H. 1975–1986. *The Lnguistic Atlas of Scotland*, 3 volumes, London: Croom Helm.
Olausson, L. and Sangster, C. 2006. *Oxford BBC Guide to Pronunciation*, Oxford: Oxford University Press.
Orton, H., Halliday, W.J., Barry, M.V., Tilling, P.M., and Wakelin, M.F. (eds.). 1962–1971. *The Survey of English Dialects*, Introduction and 4 volumes, each of 3 parts, Leeds: E.J. Arnold. (Reprinted London: Routledge, 1998.)
Ramsaran, S. 1990. RP: fact and fiction. In: *Studies in the Pronunciation of English: A Commemorative Volume in Honour of A.C. Gimson*, S. Ramsaran, 178–190, London: Routledge.
Romaine, S. 2000. *Language in Society: An Introduction to Sociolinguistics*, 2nd edition, Oxford: Oxford University Press.
Schneider, E., Kortmann, B., Burridge, K., Mesthrie, R., and Upton, C. (eds.). 2004. *A Handbook of Varieties of English: A Multimedia Reference Tool*, 2 volumes plus CD-ROM, Berlin: Mouton de Gruyter.
Smith, J. 1996. *An Historical Study of English: Function, Form and Change,* London: Routledge.
Stuart-Smith, J. 2003. The phonology of modern urban Scots. In: *The Edinburgh Companion to Scots*, J. Corbett, J.D. McClure, and J. Stuart-Smith (eds.), 110–137, Edinburgh: Edinburgh University Press.
Trudgill, P. 1999. *The Dialects of England*, 2nd edition, London: John Wiley & Sons, Ltd.
Upton, C. 1995. Mixing and fudging in Midland and southern dialects of England. In: *Studies in General and English Phonetics: Essays in Honour of Professor J.D. O'Connor*, J.W. Lewis (ed), 385–394, London: Routledge.
Upton, C. 2004. Received pronunciation. In: Schneider et al. 2004, 217–230 and also in Kortmann and Upton 2008, 237–252.
Upton, C. 2012. An evolving standard British English pronunciation model. In: *Standards of English: Codified Varieties around the World*, R. Hickey (ed.), 55–71, Cambridge: Cambridge University Press.
Upton, C., Kretzschmar Jr., W.A., and Konopka, R. 2001. *The Oxford Dictionary of Pronunciation for Current English*, Oxford: Oxford University Press.
Wells, J.C. 1982. *Accents of English*, 3 volumes, Cambridge: Cambridge University Press.
Wells, J.C. 2008. *Longman Pronunciation Dictionary*, 3rd edition, London: Longman.

15 Australian and New Zealand English

LAURIE BAUER

Introduction

Similarities and differences in settlement

Australia and New Zealand were both discovered for Anglophones by James Cook. Australia was first settled as a penal colony in 1788 and New Zealand was officially settled following the Treaty of Waitangi (signed between the Crown and the New Zealand tribes) in 1840, though by that time there was already a great deal of contact with Maori people and a fair amount of de facto settlement. There was also considerable trade by 1840 between Australia and New Zealand (Bauer 1994a: 382). With dates of settlement so close together, and the close links that have, post-British settlement, always existed between the two countries, Australia and New Zealand are often seen from a European perspective as forming a larger coherent area (the antipodes, or Australasia), losing track of the notion that the distance between Sydney and Wellington is about the same as the distance between London and Labrador.

The biggest difference between the settlements, though, is the difference in the nature of the earliest settlers. Various Australian sites were settled as places of transportation for convicts and the population was made up of those convicts and the people who were sent to be in charge of them. New Zealand was largely settled by those who desired to own land but either could not get land in Britain or had been dispossessed of the land they had held.

At later periods, there was also a big difference in the patterns of immigration, with Australia taking large numbers of settlers from Italy and Greece, and later from Vietnam, areas from which the flow of immigrants to New Zealand was relatively limited.

The Handbook of English Pronunciation, First Edition. Edited by Marnie Reed and John M. Levis.

Assumed sources

Hammarström (1980) argues, on the basis of pronunciation alone, that Australian English is derived from vernacular London English. More recent scholarship sees this as unlikely. Rather, nineteenth century London English and Australian English are mixed dialects with approximately the same inputs (Trudgill 1986; Turner 1994). The same is true for New Zealand English (Gordon et al. 2004; Hay, Maclagan, and Gordon 2008: ch. 5). The inputs are not exactly the same for Australian and New Zealand English (see above), but they are similar enough to have led to similar-sounding outputs. Where New Zealand English is concerned, the similarities are intensified by the fact that there was a great deal of contact between New South Wales and New Zealand in the second half of the nineteenth century, despite the distances involved. That high level of contact continues to this day.

The picture from pronunciation alone is not necessarily particularly clear, but when we look at other factors, the similarities between early Australian English and early New Zealand English are striking. The number of new vocabulary items they share and the number of expressions borrowed from a wide range of British dialects that they share (Bauer 1994a, 2000) can only be explained by close contact between the two.

Variation: regional, social, historical

Both Australia and New Zealand have long been said to be homogeneous linguistic areas (see Bauer 2008 for some discussion). This homogeneity is not social or ethnic, but regional. Only one regional dialect is readily recognized in New Zealand, that of southern Otago and Southland (Bartlett 1992). It is differentiated from the English spoken in the rest of New Zealand by a relatively high level of rhoticity and by a number of vocabulary items and expressions that are clearly Scots in origin. There are other regional dialects in New Zealand (Ainsworth 2004; Bauer and Bauer 2005; Kennedy 2006) but they are not part of the lay perception of dialects in New Zealand. Australia is now developing regional dialects (Bryant 1989; Bradley 1989), but they are quite new.

Social variation is readily recognized. This was addressed by Mitchell and Delbridge (1965) by assigning varieties of Australian English to one of three layers: broad, general, and cultivated. The three-way split reflects the split paraphrased by Kurath (1972: 164) in another context as one between "the folk, the middle class and the cultured". It is not clear that this three-way division ever held for Australia, given that there is considerable leakage between the levels (Bernard 1989); in New Zealand the labels were, to a large extent, adopted uncritically from the Australian experience, without any experimentation or testing. It remains a useful set of labels for dividing up the spectrum of accents in the two countries in an unsystematic way, but it cannot be given scientific content today.

There is also gender-based and ethnic variation in the two countries. I shall have little to say about these in this contribution. There is a great deal of evidence

of women leading the way in phonetic change in New Zealand as they do elsewhere in the English-speaking world (see, for example, Holmes 1997 and Maclagan 1998, 2000), and forms that appear nearer the "cultivated" end of any social spectrum are often more common in women's speech. Ethnic variation is very different in the two countries, but in New Zealand so-called "Maori English" often reflects phonetic features from the "broad" end of the social spectrum (though see Warren and Bauer 2004 for more detail and commentary on exceptions).

There is also a great deal of historical change in Australian and New Zealand English, so that recordings from the 1940s sound strange on both sides of the Tasman Sea that separates the two countries. Comments will be made on historical developments, particularly recent ones, in the course of this contribution.

Australasia as a single linguistic area with variation within it

On the basis of the discussion above, there is a sense in which we can see Australia–New Zealand as a single linguistic area with some regional diversification within it. Not only is there the evidence of lexis mentioned above, there is also a certain amount of (controversial) evidence that New Zealanders cannot recognize Australian accents as infallibly as they think they can and vice versa (Bayard 1995; Weatherall, Gallois, and Pittam 1998). Certainly, people external to the two countries have difficulty in distinguishing the two. Accordingly, in this contribution, Australia and New Zealand will be treated together. This should not be interpreted as meaning that Australian and New Zealand Englishes are "the same dialect" or "the same variety"; they are not. However, treating them together allows a relatively economical way of looking at the phonetics of the varieties.

The author's point of view

The author, though an Englishman, is resident in New Zealand and is more familiar with the New Zealand situation than with the Australian one. Accordingly, in this presentation, the New Zealand versions will tend to be taken as the default, while Australian versions are treated as variation on a New Zealand theme. This may do something to make up for the occasions where New Zealand English has been treated as a variant of Australian English. Forms that are specifically Australian or New Zealand will be marked as "AuE" or "NZE" respectively. When speaking about both together, forms will be marked as "ANZE".

Vowels

In this contribution, the individual vowels are referred to by the names of the lexical sets established by Wells (1982), except that the sets DANCE and GOLD are added to Wells' list.

Acoustics

In Table 15.1 figures for formant 1 and formant 2 of the vowels of New Zealand and Australian English are provided. The figures are derived by averaging the seven values given for New Zealand male speakers in Easton and Bauer (2000) and the two values given for Australian male speakers in the same publication. Since the earliest of these speakers were recorded 30 years before the most recent, the figures are likely to be rather conservative values.

Note that when a vowel sound is produced, two major resonators in the vocal tract, one between the lips and the point at which the tongue most obstructs the vocal tract and one between the most posterior part of the tongue and the vocal folds, each produce sound at a specific resonant frequency. These show up on a sound spectrogram or other analysis of the sound wave as bands of acoustic energy, termed formants. The relationship between formant frequency and vowel position is not always linear, but indicates relative articulatory position, with greater values for Formant 1 showing lower vowels, greater values for Formant 2 showing fronter vowels. Even in these figures, some of the shibboleths that distinguish AuE vowels from NZE vowels can be seen: the more open (lower) and retracted (backer) KIT vowel in NZ, the more open DRESS and TRAP in AuE.

Articulation, the stressed vowels

A general comment on the description of the vowels in Australian and New Zealand Englishes is probably in place. Lip rounding in these varieties is not to be equated with a pouting gesture (as it would be, for instance, in French). Rather the lips may appear quite spread and tense, but the air flow is directed through a narrow channel between the lips. The lips are held in a relatively neutral position

Table 15.1 The vowel formants (in Hz) for male New Zealand and Australian speakers.

	New Zealand English		*Australian English*	
Vowel	*F1 in Hz*	*F2 in Hz*	*F1 in Hz*	*F2 in Hz*
FLEECE	337	2296	312	2272
KIT	478	1785	373	2191
DRESS	423	2172	478	2038
TRAP	581	1956	672	1802
STRUT	736	1444	749	1362
START	767	1467	736	1318
LOT	640	1040	615	1011
THOUGHT	414	815	438	791
FOOT	455	1106	408	881
GOOSE	371	1654	362	1651
NURSE	426	1734	489	1513

at virtually all times. A "rounded" vowel in these varieties is thus not articulatorily the same as a rounded vowel in RP, and this may have an effect (yet to be determined) on the acoustic characterization of the vowels. For more details on vowels in general see Bauer and Warren (2004).

FLEECE Although listed as a monophthongal vowel in Tables 15.2 and 15.3 (see later sections), FLEECE is frequently diphthongized, especially in Aus, with a short lower onglide.

KIT This vowel, as it appears in stressed syllables, is the main shibboleth distinguishing AuE and NZE varieties. Australians accuse New Zealanders of saying *fush and chups*, while New Zealanders accuse Australians of saying *feesh and cheeps*. Neither side is right at the phonological level, though phonetically it is true that the Australian vowel is close and front, while the New Zealand vowel is much more open and centralized. How open the NZE variant is depends partly upon context and partly upon social factors. The transcription for this vowel used here, /ɘ/, represents a variant from nearer the cultivated end of the social spectrum, with more open variants being, in general terms, broader variants. The most open variants reach a position between [ə] and [ɐ]. In NZE there is a closer variant of this vowel found before [ŋ]. It may be that this variant should be assigned to a different lexical set: HAPPY or FLEECE (despite being relatively short).

DRESS The DRESS vowel is more open in AuE (where it is typically transcribed as [ɛ], although that makes it look rather more open than it actually is) than in NZE (where it is typically transcribed as [e], which also makes it look more open than it actually is). Innovative NZE pronunciation in the last few years has found some tokens of this vowel overlapping with tokens of FLEECE (Bell 1997), from which it is distinguished in terms of vowel length. Despite this, DRESS is quite a long vowel in both AuE and NZ, which helps explain the neutralization with SQUARE in NZE before /r/ (see the next section below on neutralization).

TRAP One of the biggest changes in AuE in the last thirty years or so is the move to a much more open pronunciation of TRAP. This follows chronologically, if not causally, the similar move in standard varieties of British English. The vowel remains no further back than central. This contrasts with NZ, where the close variety that used to be typical of advanced RP, vernacular London speech, and Australian speech as well, is still heard. This is typically pronounced close to open-mid [ɛ], and is correspondingly mistaken for the DRESS vowel by speakers from Britain, North America, and even from Australia.

DANCE This vowel is added to the list of vowels in Wells' (1982) set because of its sociolinguistic importance in AuE. Although there used to be some variation in the DANCE set in NZ, with the TRAP and the START vowel both being heard, that has vanished within living memory, and only the START vowel is now used. In AuE the

situation is far more complex, with the choice of realization of the DANCE set reflecting lexical, social, and regional choices (Bradley 1989: 263; Cox and Palethorpe 2000: 40).

STRUT The STRUT vowel is very open and very front, typically rather advanced from central in modern usage. Its length is variable, which will be discussed in the section below on articulation.

START The START vowel overlaps with the STRUT vowel in quality, so that they can be taken to be long and short members of a pair of vowels with the same quality. As was noted for STRUT, this vowel is very open and typically advanced from the central position. It is consistently long, but not diphthongized.

LOT The LOT vowel is back, open, and rounded, but not peripheral (that is, not pronounced near the perimeter of the vowel space) being both centralized and raised from the position assigned this symbol by the IPA. Like the vowel in systems of south-east England, it is rounded, but rather closer and more centralized than is typically described for RP. There are occasional traces of a distinct CLOTH vowel from older speakers (with phonemically the same quality as THOUGHT), but this is now sporadic and largely idiosyncratic.

THOUGHT The THOUGHT vowel is much closer than the corresponding vowel in RP and the transcription as [o] is accurate in indicating the vowel height (with occasional NZE tokens even closer), but the vowel is not peripheral. The vowel is consistently long and may be diphthongized in lengthening environments, especially in the phrase-final position where it may, in New Zealand, become disyllabic (compare NEAR and SQUARE below).

FOOT Until recently, the FOOT vowel could be seen as the short congener of THOUGHT in the system (see, for example, Maclagan 1982). The last 15–20 years have seen quite considerable evolution in the quality of this vowel, in line with developments in British and North American Englishes. The vowel has long been unrounded in the expression *gidday* (a greeting) as the <i> spelling attests, but that unrounded and fronted pronunciation has escaped from that fixed expression and is now used not only in the word *good* in isolation but in general for the FOOT vowel. I know of no acoustic studies of the new vowel quality, but auditorily it gives the impression of being as far forward as the GOOSE vowel, so that to use [ɨ] to transcribe it is to show it as being further back than it really is. Lay speakers are not aware of this development.

GOOSE As in many other varieties of English, GOOSE has become a front vowel. It should probably be transcribed as [ʏ], but it is not close enough for [y], as can be heard in some Scottish varieties. Especially in AuE it is often diphthongized, with a lower onglide, but less so than FLEECE, both in terms of the degree of onglide and in terms of the frequency with which the onglide occurs. In NZE and some regional

varieties of AuE, it remains back before historical /l/, with the /l/ itself often elided completely (see the subsection below on /l/).

Nurse The NURSE vowel is monophthongal, front of central and relatively close, and in NZE is rounded so that [ɵː] is a reasonable transcription. This means that in NZE there is overlap with the GOOSE vowel.

Near, square, cure These vowels are very different in AuE and NZE. In most Australian varieties, they are all long monophthongs (though Bradley 1989: 264 points to considerable regional variation), especially but not exclusively before /r/. In NZE there has been a long period of increasing merger between NEAR and SQUARE (resulting in a number of new homophones such as *beer*, *bare*, *pier*, *pear*, *hear*, *hair*, *really*, *rarely*, and so on). Many young NZE speakers cannot produce a difference, and some cannot hear a difference (Hay, Warren, and Drager 2006) . For such speakers the onset of the vowel varies between [i] and [e], but since phonemic /e/ is often pronounced with an [i]-like resonance, the distinction can be hard to hear in isolation.

In NZE, both the NEAR and SQUARE vowels (or all three for those who maintain a NEAR-SQUARE distinction) are diphthongs or disyllabic sequences (especially in lengthening environments). The first elements may be transcribed as [i], [e], and [ʉ] (or [ʏ], corresponding to a variation in GOOSE). The second element in the diphthong is a very open central vowel, which may be transcribed as [ɐ].

There is some variation, as there is in RP and other British varieties, between CURE and FORCE in words like *moor*, *tour* – the FORCE variant is rather less used in ANZE than in British varieties; *fewer* (if used at all) has CURE, often with two syllables, rather than FORCE.

Face, price, choice Using the terminology from Wells (1982), there is diphthong shift between these vowels in RP and the way they are realized in Australian and New Zealand English. Therefore, despite "cultivated" variants, which are only slightly displaced from the corresponding RP vowels, the onset in FACE is typically more open, the onset in PRICE is typically backer, and in NZE the onset in CHOICE is typically closer than the corresponding realizations in RP, while in AuE it seems the onset is becoming more open (Cox and Palethorpe 2000). The onset in PRICE is usually unrounded, but may be rounded in broader variants. The second element may be transcribed [ɪ] in Aus, where the KIT vowel is close, but has to be transcribed as [i] in NZ, where the KIT vowel is often very open. In both varieties, a transcription with [e] may be more realistic.

Mouth There is considerable variation in the quality of the onset to the diphthong in MOUTH, with more [ɛ]-like variants belonging to the broader end of the social spectrum. There is also some loss of rounding of the second element, some centralization of the second element, and even monophthongization, with a stereotypical version of *Now is the Hour* (the title of a famous Maori song) as *Nar is the are*. I know of no investigation of the social implications of the various versions of this vowel. It might seem that

the monophthongized version would be the variant to call forth /r/ sandhi, but it appears that "intrusive" /r/ can appear with any variant of this phoneme.

GOAT The GOAT vowel is diphthongal with a very open, central first element, and no rounding.

GOLD Because of the effects of /l/-vocalization, there can be a phonemically separate GOLD vowel in NZ, in words like *coal*, which can contrast with *Coe*, or in *gold*, contrasting with *goad*. The NZE GOLD vowel has a more rounded first element than the GOAT vowel.

Neutralization

The patterns of neutralization of stressed vowels before /r/ and /l/ are not the same in AuE and NZE. NZE has a fuller set of these neutralizations, especially before /l/, while the patterns of neutralization in AuE appear to be still developing. The difficulty in describing these neutralizations is that they are sociolinguistically variable, with the result that neutralization is not always a transitive relationship: that is, if A is neutralized with B and B is neutralized with C in the same environment, it does not follow for any given speaker that A will be neutralized with C. Table 15.2 sets out the most developed cases of neutralization, with examples and comments.

Articulation, the unstressed vowels

There are no COMMA-KIT minimal pairs in unstressed syllables: *villages* and *villagers*, *chatted* and *chattered* are either homophonous or are consciously distinguished by the use of the NURSE vowel in *villagers* and *chattered*.

This means that there are basically two unstressed vowels to be considered, with a third one arising from the vocalization of /l/. The happy vowel is phonemically associated with the FLEECE vowel, the two vowels of *seedy* being perceived as "the same" while the two of *city* are clearly different in NZE. This is supported by the fact that the happy vowel can be diphthongized, as the FLEECE vowel may be, though whether the range of diphthongization is the same in the two cases has not been investigated. The happy vowel is also often rather longer than might be expected, but this will be considered in the next section below on general comments on length.

The other vowel, which we can term the commA vowel, has a range of realizations that may overlap auditorily with STRUT, especially in the word-final position or with NZE KIT. In the phrase-final position this vowel is often lengthened considerably (see also below in the next section).

At least in NZ, this vowel is typically used in contexts where RP would have syllabic consonants, providing a vocalic nucleus for these weak syllables. This vowel is also used to distinguish a few pairs such as *groan* and *grown*, where the past participle marker has a full syllable with the comma vowel.

Table 15.2 Cases of neutralization.

Vowels involved	*Environment*	*Examples*	*Comments*
DRESS-SQUARE	/__r	*ferry* and *fairy* become homophonous	General in NZ, not Aus
GOOSE-CURE	/__r	A word like *fluoride* may contain either vowel	
DRESS-TRAP	/__l	*telly* and *tally* become homophonous	Ubiquitous in NZ, widespread in some regions of AuE. In NZE the output may be perceived as a token of MOUTH, despite the fact that MOUTH is phonetically distinct, so that *twelve* may be considered as belonging to the MOUTH set
FOOT-GOOSE	/__l	*pull* and *pool* become homophonous	Ubiquitous in NZ, only in South Australia in Australia
LOT-GOAT	/__l	*doll* and *dole* become homophonous	Ubiquitous in NZ
LOT-STRUT	/__l	*cult* and *colt* may become homophonous	*colt* may contain the GOLD vowel rather than LOT, esp. if the /l/ is elided
FLEECE-NEAR	/__l and /__r	*reel* and *real* are homophonous; a word like *fearing* may contain either vowel	*really* is pronounced with a number of vowels in different regions and styles in AuE

General comments on length

As in the south-eastern varieties of British English from which ANZE fundamentally derives (neither AuE nor NZE retain phonological traces of distinctively northern or south-western British features), there is a distinction between long and short vowels. Despite the distinction being fundamentally the same as that found in Britain, the actual vowels that can be paired in terms of length are not the ones that are paired by the Jones' systems of transcription, nor those paired by the orthographic system. Phonetic pairings are those shown in Table 15.3. It should

Table 15.3 Pairings of vowels by length.

Long vowel	*Corresponding short vowel AuE*	*Corresponding short vowel NZE*	*Comments*
FLEECE	KIT	DRESS	
START	STRUT	STRUT	
THOUGHT	FOOT	FOOT	This is now a slightly old-fashioned pronunciation, see comments on FOOT in the text
GOOSE	–	–	
NURSE	COMMA	KIT	These matches are not always accurate: NURSE is often rounded and KIT may be very open in NZ, and COMMA may be very open in AuE
SQUARE	DRESS	–	SQUARE is monophthongal in Aus

also be noted that the STRUT vowel is not strictly a checked vowel, as it is in RP. The common phrase *see ya!* (a farewell) regularly ends in a stressed STRUT vowel.

As well as this phonological vowel length, there are two types of phonetic vowel length. The first is the type also found in other varieties of English, whereby vowels are lengthened in syllables without codas or in syllables where the coda is a voiced obstruent. Such lengthening has been referred to above as occurring "in lengthening environments". Nothing further will be said about this. There is also important prosodic lengthening, particularly at phrase boundaries or for emphasis. In most places, this does not disturb phonological structure, but a phrase-final STRUT may be perceived by speakers of these varieties as being the START vowel. It should be noted that this is true even when the STRUT vowel in question is the COMMA vowel, which has become more open in the phrase-final position. Thus a phrase like *Look at the koala* can be perceived as having final START. This same phenomenon may account for the final HAPPY vowel being perceived as being the same as the FLEECE vowel.

Consonants

There is little in the consonantal system of Australasian English that is surprising when it is compared with Northern Hemisphere varieties. The loss of /r/ from words like *far* and *farm*, and the subsequent and on-going loss of /l/ from words

like *fill* and *film*, are familiar from many British varieties, while /j/-dropping (which gives /nuː/ for *new*) is familiar in more advanced forms from some British dialects and from North American varieties. Even the variation in plosives is not greatly different from that found elsewhere.

Plosives

The voiced plosives /b/, /d/, and /g/ are weakly voiced, as in other varieties of English. Where they are in coda-position, the length of the preceding sonorant is often the main clue available as to the phonological voicing of the plosive, sonorants being longer before phonologically voiced obstruents than before voiceless ones.

The grave (i.e., noncoronal) voiceless plosives /p/ and /k/ behave differently from the coronal (more narrowly, alveolar) /t/. Initially in a stressed syllable, all of these plosives are aspirated/affricated. The plosive /p/ is usually just aspirated, the other two are usually affricated to a greater or lesser degree, the quality of the friction occurring after /t/ suggesting a tongue-tip articulation for the plosive. Intervocalically before an unstressed or weakly stressed syllable, /p/ and /k/ are aspirated (/k/ is probably affricated), while /t/ is voiced (Silby 2008), usually with a quick enough articulation for a transcription as [ɾ] to be reasonable, but sometimes with an articulation that is not easily distinguishable from a [d]. In the pre-consonantal or pre-pausal position, the voiceless plosives may be unreleased, weakly released, aspirated (affricated), or glottalized (as set out in Table 15.4). Research on the distribution (phonological or sociological) or these variants (Holmes 1995a, 1995b) is now outdated, and incomplete in that it covers only /t/. Where glottalization is employed, there may be either glottal reinforcement or glottal replacement (see Wells 1982 for the terminology). Where glottal replacement is found (as in the utterance overheard between NZ students a few years ago of [ʃɐːʔ ɐːːʔ] "shut up"), the glottal stop may take on phonemic status in its own right. The glottalized variants are far more current in NZE than in AuE (see Tollfree 2001 on the situation in Australian English), and have arisen in the last fifteen years or so. This material is summarized in Table 15.4.

Table 15.4 Allophones of voiceless plosives.

	Bilabial	*Velar*	*Alveolar*
$[_{ft}$ ___	p^h	k^x	t^s
	pan	can	tan
V ___ V	p^h	k^x	ɾ
	seeping	seeking	seating
___ #	p, p^h, $\widehat{ʔp}$, ʔ (NZ)	k, k^x, $\widehat{ʔk}$, ʔ	t, t^s, $\widehat{ʔt}$, ʔ
	lop	lock	lot
___ C	p, $\widehat{ʔp}$, ʔ	k, $\widehat{ʔk}$, ʔ	t, $\widehat{ʔt}$, ʔ
	lops	locks	lots

Fricatives

As in other varieties of English, the voiced fricatives are only weakly voiced, as also noted above for the voiced plosives. In NZE this sometimes results in some apparent movement between voiced and voiceless categories: for instance, *president* and *precedent* may become homophonous, the first fricative in *positive* may be variably voiced or voiceless, as may the first fricative in *pessimistic*. *Equation* has /ʃ/ rather than /ʒ/, which makes it morphophonemically regular. While *thither* usually has a voiceless initial fricative, this is presumably due to a voiceless model from Scottish English, rather than a local devoicing. In Maori English, /z/ (particularly in the final position, e.g., in *freeze*) is said to be devoiced even more than in mainstream New Zealand English (Bell 2000; Warren and Bauer 2004).

The fricatives /θ/ and/ð/ are increasingly replaced by /f/ and/v/ in the speech of younger speakers. Some words, such as *with*, are particularly strongly affected (Campbell and Gordon 1996; Wood 2003). This is not perceived as standard at this stage, however. In Maori English these fricatives may be affricated as [tθ] and [dð] (see Bell 2000).

The fricative /s/ is replaced by /ʃ/ (a) when followed by a phonemic /j/, where the /j/ may or may not be assimilated, so that *consume* can be [kənʃʉːm] or [kənʃjʉːm]; (b) optionally before a /ʧ/, as in *student* /ʃʧʉːdənt/; and (c) increasingly before /tr/, as in *strong* [ʃtɹɒŋ]. These forms are not proscribed and those in (a) may be considered cultivated. The voiced equivalent /z/ is palatalized only in environments corresponding to (a), so that *presume* can be [pɹɘʒʉːm] or [pɹɘʒjʉːm].

/l/

The phoneme /l/ is typically pronounced with a darker quality than in British RP, though for many speakers there is nonetheless a difference in quality between pre-vocalic and pre-consonantal allophones. However, there is variable vocalization of /l/ pre-consonantally or finally. In parallel with the historically earlier vocalization of /r/, this leads to instances of linking /l/, so that there may be no phonetic [l] in *feel bad* ([fiːɯ bæd]) but a phonetic [l] in *feel it* (fiːlʲ ɪt]).

The vocalization of /l/ leads to a range of phonetic outputs. In the simplest cases there is a back vowel, typically unrounded, of variable vowel height, perhaps [ɯ] or [ɤ]. In some cases the /l/ appears not to gain any realization at all (see just below) and in yet other cases, the /l/ is realized as a lengthening of the previous vowel (in some cases accompanied by a distinctive quality of that vowel). The clearest example of this last phenomenon is restricted to New Zealand and happens after the GOOSE vowel (possibly neutralized with the FOOT vowel before a historical /l/ anyway). In isolation, the GOOSE vowel is very front (transcribed here as [ʉ], but occasionally further forward than that suggests, possibly [ʏ]); where there is a following underlying /l/, this vowel is realized as [u]. In a word like *school* or *pool*, this [u] may be lengthened, but remain a monophthong: [skuː]. Similarly, in a word like *milk*, with the KIT vowel followed by /l/ and an obstruent, the /l/ may not be present, but the preceding vowel may be realized as [ʏ], [mɤk] (NZE), although this is only one of a number of potential realizations of this word, others including [miɤk] or [miɯk]

(with a close onset to the diphthong). Following long back vowels as in *walls* there may be only minimal or no diphthongization of the vowel, and so no marker of the erstwhile /l/ at all, with pronunciations like [woːz] or [woᵊz].

Before /l/ there is considerable neutralization of vowel contrasts, which was discussed above in the section on neutralization.

/ɹ/

Both Australian and New Zealand Englishes are typically described as nonrhotic, but this hides a multitude of variable usages. Not only is there the Southland-Otago accent of New Zealand, which is typically characterized as rhotic, even though it tends to be rhotic only in the context of an immediately preceding NURSE vowel, as in *word* [wɵːrd] with some variable rhoticity following the LETTER vowel (letter may be [lerəɹ]) and considerably less, if any, rhoticity following FORCE and START (in words like *warm* and *farm*), but there is also increasing rhoticity on both sides of the Tasman, particularly after NURSE and LETTER, but also occasionally elsewhere, in a way that at the moment appears to be essentially random. In New Zealand, this increasing rhoticity appears to be partly dialectal and partly ethnic (greater among Maori and Pasifika speakers – i.e., speakers with Pacific Island ethnicities), but is heard sporadically from speakers who do not belong to these groups (Kennedy 2006: Marsden 2013). Surprisingly, linking rhoticity in NZE has spread to following the MOUTH vowel, so that *how/r/ever* and *now/r/ and again* are frequently heard (Hay and Warren 2002).

The /r/ in all these instances is an apical alveolar-to-post-alveolar approximant [ɹ]. Devoicing and frication of the /r/ are found as in other varieties of English, following voiceless plosives and alveolar plosives respectively. An allophone [ɾ] is variably heard following /θ/ in words like *through*, although [ɾ] in the intervocalic position would be perceived as realizing an alveolar plosive.

The semi-vowels [j] and [w]

As in other varieties of English, [j] and [w] are devoiced and may be fricated following an initial voiceless plosive in a stressed syllable, so that *pewter* and *cute* may be [pçuːtɐ], [pʲjuːtɐ], [kˌçuːt], [kˌʲjuːt]. Where an alveolar plosive and [j] arise in a cluster, the output is generally an affricate, [ʧ] or [ʤ], so that *dune* and *June* become homophones.

Where one of FLEECE, GOOSE, GOAT, FACE, PRICE, CHOICE vowels (but not NZE MOUTH, for which see above) forms a sequence with another vowel, a glide arises between the two vowels, agreeing in backness and rounding with the first vowel, to prevent the hiatus. This occurs in sequences like *see*[j] *it*, *be*[j] *in*, *do*[w] *it*, *lie*[j] *in*, *de*[j]*ontic*, *go*[w] *on*, *pro*[w]*active*, and so on. These intrusive elements are distinct from the full phonemes /j/ and /w/: *say 'S'* is not homophonous with *say 'yes'*, nor is *know it* homophonous with *no wit*. The intrusive elements are shorter, less firmly articulated and (where [w] is concerned) often less rounded; they are nevertheless auditorily distinguishable from the vowels that surround them.

Yod-dropping (Wells 1982), the loss of /j/ immediately following a coronal consonant, is variable. It has vanished completely following /s/ (as in

superintendant) in the last 20 years, is still variable following /θ/ (as in *enthusiasm*), and is apparently partly lexically determined following /n/ (as in *new*; in New Zealand the loss is particularly noticeable in the item *New Zealand*). Following /t/ or /d/ the result is an affricate rather than /j/-loss, as described above.

In conservative varieties (and to a certain extent, in southern accents of New Zealand), /hw/ in *which* is still distinct from /w/ in *witch*, but the distinction is dying fast.

Nasals

There is little to say about the nasal consonants themselves, which behave as in other major varieties of English, but it should be noted that the nasality from nasal consonants easily spreads to adjacent segments. This is dealt with further below in the section on voice quality.

Prosodics

Stress

Word stress functions basically as in other standard "inner circle" varieties of English, and while there are minor lexical differences in occurrence, these do not disturb the fundamental system. However, even speakers who use this system of stress seem unsure about it. I was in the interesting position recently of having a class of first-year undergraduates tell me that *revenue* is stressed on the third syllable, even though they were pronouncing it with first-syllable stress. Moreover, in broadcasting, stress is more variable than might be expected, perhaps particularly so in noun–noun constructions, where the position of the main stress is notoriously difficult to predict.

Rhythm

While the fundamental underlying stress-timing inherited from British English is still present, it appears to be weakening, probably more so in NZE than in AuE. In the first place, this is due to the use of full vowels (the phonetic nature of which is determined by spelling pronunciation) where unstressed vowels are normal in RP or other British and American varieties. Pairs such as *effect* and *affect*, *Johnston* and *Johnstone*, are often distinguished by the use of different full vowels (recall that there is no COMMA-KIT distinction in unstressed syllables; speakers who cannot distinguish *villagers* from *villages* equally cannot use these means to distinguish *effect* and *affect*; see Bauer 1994b). Even grammatical words may be heard with full vowels in contexts where they are not stressed. This has the effect of leveling out the difference between stressed and unstressed syllables. In New Zealand, the trend away from stress-timing may be exaggerated by the effect of Maori English, where the rhythm is based on the original mora-timing of Maori, now being lost as vowel length is eroded in Maori (Maclagan et al. 2004).

Intonation

Discussion of the intonation of Australian and New Zealand Englishes has tended to focus on the High Rise Terminal first noted in print by Benton (1965) and discussed in detail for Australian English in Guy et al. (1986). This rise occurs on statements, and is used as a pragmatic device to check comprehension or to draw attention to critical parts of a narrative (Warren and Britain 2000; Warren 2005). It is perceived by outsiders as a questioning intonation, but is phonetically distinct from the intonation used to ask real questions (Warren and Daly 2005).

Otherwise the intonation patterns of these varieties are, in everyday usage, rather flat, and not as varied as RP is reported to be. Ainsworth (2004) reports on more varied intonation patterns in one area of New Zealand.

Voice quality

It seems likely that one of the distinctive features of English from this part of the world is voice quality, and it may also be that voice quality helps distinguish ethnic varieties in New Zealand and social varieties everywhere. However, this area has not been investigated in any depth or in any phonetic detail. Features that seem to be relevant include a generally relaxed articulation, including lack of great articulatory precision and for some varieties a rather slow delivery (though see Robb, Maclagan, and Chen 2004 on the speed of NZE), an overall back resonance, and variable nasalization. The nasalization varies from the effect of adjacent nasal consonants to widespread nasal air flow, so that the vowels in *had* and *ham* are not auditorily distinct in their nasal quality.

Conclusion

This survey should have indicated that there are considerable differences between AuE and NZE, despite there being many similarities between them. The unity arises from the fundamental phonological structure of the systems, inherited from southeastern varieties of British English. Although both AuE and NZE have other influences operating upon them – Irish and Scottish varieties of English, contact languages and the like – they can be considered to be English varieties which have undergone phonetic change, and in a few cases that phonetic change has led to phonological differences both from the input and from the varieties now heard in Britain.

REFERENCES

Ainsworth, H. 2004. *Regional variation in New Zealand English: the Taranaki singsong accent*. Unpublished PhD thesis, Victoria University of Wellington.

Bartlett, C. 1992. Regional variation in New Zealand English: the case of Southland. *New Zealand English Newsletter* 6: 5–15.

Bauer, L. 1994a. English in New Zealand. In: *The Cambridge History of the English Language, vol. V: English in Britain and Overseas, Origins and Developments*, R.W. Burchfield (ed.), 382–429, Cambridge: Cambridge University Press.

Bauer, L. 1994b. Spelling pronunciation and related matters in New Zealand English. In: *Studies in General and English Phonetics*. J.W. Lewis (ed.), 320–325, London: Routledge.

Bauer, L. 2000. The dialectal origins of New Zealand English. In: *New Zealand English*, A. Bell and K. Kuiper (eds.), 40–52, Amsterdam/Philadelphia: Benjamins and Wellington: Victoria University Press.

Bauer, L. 2008. Homogeneity, heterogeneity and New Zealand English. *New Zealand English Journal* 22: 1–8.

Bauer, L. and Bauer, W. 2005. Regional dialects in New Zealand children's playground vocabulary. In: *Languages of New Zealand*, A. Bell, R. Harlow and D. Starks (eds.), 194–216, Wellington: Victoria University Press.

Bauer, L. and Warren, P. 2004. New Zealand English: phonology. In: *A Handbook of Varieties of English*, B. Kortmann, E.W. Schneider, K. Burridge, R. Mesthrie, and C. Upton (eds.), 580–602,Berlin and New York: Mouton de Gruyter,

Bayard, D. 1995. *Kiwitalk*, Palmerston North: Dunmore.

Bell, A. 1997. Those short front vowels. *New Zealand English Journal* 11: 3–13.

Bell, A. 2000. Maori and Pakeha English: a case study. In: *New Zealand English*, A. Bell and K. Kuiper (eds.), 221–248, Amsterdam/Philadelphia: Benjamins and Wellington: Victoria University Press.

Benton, R. 1965. *Research into the English Language Difficulties of Maori School Children, 1963–1964*. Wellington: Maori Education Foundation.

Bernard, J.R. 1989. Quantitative aspects of the sounds of Australian English. In: *Australian English*, P. Collins and D. Blair (eds.), 187–204, St. Lucia: University of Queensland Press.

Bradley, D. 1989. Regional dialects in Australian English phonology. In: *Australian English*, P. Collins and D. Blair (eds.), 260–270, St. Lucia: University of Queensland Press.

Bryant, P. 1989. Regional variation in the Australian English lexicon. In: *Australian English*, P. Collins and D. Blair (eds,), 301–314, St. Lucia: University of Queensland Press.

Campbell, E. and Gordon, E. 1996. "What do you fink?" Is New Zealand English losing its "th"? *New Zealand English Journal* 10: 40–46.

Cox, F. and Palethorp, S. 2000. The changing face of Australian English vowels. In: *English in Australia*, D. Blair and P. Collins (eds), 17–44, Amsterdam: Benjamins

Easton, A. and Bauer, L. 2000. An acoustic study of the vowels of New Zealand English. *Australian Journal of Linguistics* 20: 93–117.

Gordon, E., Campbell, L., Hay, J., Maclagan, Sudbury, A, and Trudgill, P. 2004. *New Zealand English: Its Origins and Evolution*, Cambridge: Cambridge University Press.

Guy, G., Horvath, B., Vonwiller, J., Daisley, E., and Rogers, I. 1986. An intonational change in progress in Australian English. *Language in Society* 15: 23–51. doi: 10.1017/S0047404500011635.

Hammarström, G. 1980. *Australian English: Its Origin and Status*, Hamburg: Buske.

Hay, J., Maclagan, M., and Gordon, E. 2008. *New Zealand English*, Edinburgh: Edinburgh University Press.

Hay, J. and Warren, P. 2002. Experiments on /r/-intrusion. *Wellington Working Papers in Linguistics* 14: 47–58.

Hay, J., Warren, P., and Drager, K. 2006. Factors influencing speech perception in the context of a merger-in-progress *Journal of Phonetics* 34(4): 458–484.

Holmes, J. 1995a. Time for /t/: initial /t/ in New Zealand English. *Australian Journal of Linguistics* 15: 127–156.

Holmes, J. 1995b. Two for /t/: flapping and glottal stops in New Zealand English. *Te Reo* 38: 53–72.

Holmes, J. 1997. Setting new standards: sound changes and gender in New Zealand English. *English World-Wide* 18: 107–142.

Kennedy, M. 2006. Variation in the pronunciation of English by New Zealand school children. Unpublished MA thesis, Victoria University of Wellington.

Kurath, H. 1972. *Studies in Area Linguistics*, Bloomington: Indiana University Press.

Maclagan, M.A. 1982. An acoustic study of New Zealand vowels. *The New Zealand Speech Therapists' Journal* 37: 20–26.

Maclagan, M.A. 1998. Women and language change in NZE: the case for considering individual as well as group data. *Te Reo* 41: 69–79.

Maclagan, M.A. 2000. How long have women been leading language change? In: *Gendered Speech in Social Context: Perspectives from Town and Gown*, J. Holmes (ed.), 87–98, Wellington: Victoria University Press.

Maclagan, M.A., Harlow, R., King, J., Keegan, P., and Watson, C. 2004. New Zealand English influence on Maori pronunciation over time. *Te Reo* 47: 45–57.

Marsden, S. 2013. Phonological variation and the construction of regional identities in New Zealand English. Unpublished PhD thesis, Victoria University of Wellington.

Mitchell, A.G. and Delbridge, A. 1965. *The Pronunciation of English in Australia*, Revised edition. Sydney: Angus and Robertson.

Robb, M.P., Maclagan, M.A., and Chen, Y. 2004. Speaking rates of American and New Zealand varieties of English. *Clinical Linguistics and Phonetics* 18(1): 1–15.

Silby, W. 2008. t-Flapping in New Zealand English: a change over time. *New Zealand English Journal* 22: 24–35.

Tollfree, L. 2001. Variation and change in Australian English consonants. In: *English in Australia*, D. Balir and P. Collins (eds.), 45–67, Amsterdam: Benjamins.

Trudgill, P. 1986. *Dialects in Contact*, Oxford: Blackwell.

Turner, G.W. 1994. English in Australia. In *The Cambridge History of the English Language, vol. V: English in Britain and Overseas, Origins and Developments*, R.W. Burchfield (ed.), 277–327, Cambridge: Cambridge University Press.

Warren, P. 2005. Patterns of late rising in New Zealand – intonational variation or intonational change? *Language Variation and Change* 17(2).

Warren, P. and Bauer, L. 2004. Maori English: phonology. In: *A Handbook of Varieties of English*, B. Kortmann, E.W. Schneider, K. Burridge, R. Mesthrie, and C. Upton (eds.), 614–624, Berlin and New York: Mouton de Gruyter.

Warren, P. and Britain, D. 2000. Prosody and intonation of New Zealand English. In: *New Zealand English*, A. Bell and K. Kuiper (eds.), 146–172, Wellington: Victoria University Press.

Warren, P. and Daly, N. 2005. Characterising New Zealand English intonation – broad and narrow analysis. In: *Languages of New Zealand*, A. Bell, R. Harlow, and D. Starks (eds.), 217–237, Wellington: Victoria University of Wellington Press.

Wells, J.C. 1982. *Accents of English*, 3 vols, Cambridge: Cambridge University Press.

Weatherall, A., Gallois, C., and Pittam, J. 1998. Australasian attitudes towards Australasian accents. *Te Reo* 41:153–162.

Wood, E. 2003. Th-fronting: the substitution of f/v for θ/ð in New Zealand English. *New Zealand English Journal* 17: 50–56.

16 The Pronunciation of English in South Africa

IAN BEKKER AND BERTUS VAN ROOY

Introduction

This article offers a brief but general overview of the evolution of South African English (SAfE) as well as its current characteristics, both from a descriptive point of view as well as from the point of view of what might be referred to as the "social life" of this dialect, i.e., the linguistic system's diachronic and synchronic relationships with social factors and forces. In line with the volume of which this chapter forms a part, emphasis will fall on the pronunciation features of SAfE. Details will be provided not only for the standard variety (General White SAfE) but also the various sociolects (e.g., Broad SAfE), ethnolects (e.g., South African Indian English) as well as L2 varieties (especially Black South African English, the numerically strongest and best researched of the nonancestral SAfE dialects).

In what follows, the social history of SAfE is first sketched, detailing its emergence via a complex nineteenth century koineization process and then focusing on subsequent developments. The process of the transmission of English to nonancestral communities also receives attention. The next section then provides an overview of the various varieties' pronunciation features while the chapter ends with a section overviewing current developments in the field and a conclusion.

The historical sociolinguistics of South African English

The history of English in South Africa begins with the first British occupation of the Cape in 1795 (Giliomee and Mbenga 2007: 85). On the standard account it is not, however, until the arrival of the 1820 settlers in the Eastern Cape (see Figure 16.1) that a new dialect of English is born.

This episode in the colonial history of South Africa constituted what Trudgill (2004: 26) refers to as a "tabula rasa" context, i.e., "those in which there is no prior-existing population speaking the language in question, either in the location or

The Handbook of English Pronunciation, First Edition. Edited by Marnie Reed and John M. Levis.

Figure 16.1 A map of South Africa

nearby". In other words, koineization took place among the various English dialects that served as inputs, the output of which was a new variety of English, which has been referred to as Cape English (CpE) in, for example, Bekker (2012). The standard picture, for example, in Lanham and Macdonald (1979) or Lass (1995), is that the 1820 Settlers were mainly of lower class origin and predominantly from the south-eastern part of England (including London). The (over)simplistic picture is, therefore, of a CpE reflecting many of the trends of early nineteenth century Cockney (and similar in many linguistic respects therefore to Australian English). However, according to contemporary historians such as Welsh (1998: 127) and Gilliomee and Mbengwa (2007: 85–86), the eventual group of about 4000 settlers, who were selected from among approximately 80 000 applicants, included a higher proportion of middle class, educated settlers, many of whom had some means upon their arrival in the Cape, and who did not intend to become farmers or laborers. The received view among linguists of the predominantly lower class origin of the settlers is thus challenged by historians, and an updated view may help to explain why SAfE, unlike the other Southern Hemisphere varieties, does

not display some typical Cockney features, e.g., the use of *-in* for *-ing* for the present participle (*talkin'* for *talking*). Another complication lies in the fact that the settlement area had already been populated to a degree by Cape-Dutch/Afrikaans speakers. There was much intensive contact (e.g., intermarriage) between the English and Afrikaans groups, even if political relations were often strained (Branford 1996: 38–39) and there is some debate in the literature as to whether SAfE (and thus by implication CpE) was influenced by Afrikaans on more than just a superficial level (i.e., on a structural as opposed to purely lexical or lexicogrammatical level), with Lanham and Macdonald (1979), Jeffery and van Rooy (2004), and Wasserman (2014), for example, supporting the notion, while Lass and Wright (1986) and Mesthrie (2002a) argue against it.

The second phase in the formation of SAfE was a period of settlement during the 1840s to 1850s and focused on Natal (now KwaZulu-Natal – see Figure 16.1). Here the standard picture is that the relevant settlers were of a middle to upper class origin, that there was virtually no Afrikaans influence on the koineization process and that there was a distinctly North of England bias, although this bias was no doubt tempered, although not completely, by the use of Standard English (and thus an early form of Received Pronunciation) by many of these middle class to upper class individuals. The output of the koineization process can usefully be termed Natal English (NE) and for many commentators the formation of SAfE ends here. This standard model of the formation of SAfE is, for example, made explicit in Schneider (2007: 176) who explains, with respect to the Eastern Cape and Natal periods, that "in both cases a recognizable founder effect is worth noticing: despite their relatively small numbers … these two groups laid the foundations for the main accents of present-day SAfE". Bekker (2012), however, argues that an important third phase took place during the birth and development of Johannesburg, which was itself based on the discovery of gold on the Witwatersrand. A discussion of the technical details is not appropriate here, but in essence the argument is that Johannesburg constituted yet another tabula rasa context (Trudgill 2004: 26) and that a third koineization process took place, inputs into which included CpE, NE, and a whole gamut of other English accents (British as well as colonial, e.g., Australian and American) as well as L2 varieties such as the English spoken by L1 Afrikaans and L1 Yiddish speakers. With respect to the later, the immigrants to early Johannesburg included a sizeable number of mainly Eastern European Jews, particularly from Lithuania and Latvia (Kaplan and Robertson 1991). This group became the first nonancestral population to become fully incorporated into the White SAfE speech community.

As argued by Bekker (2012), the output of this last and third koineization process was a sociolectal continuum that many refer to as "South African English", i.e., that variety still spoken primarily (although certainly not exclusively) by white L1 speakers of English in South Africa and henceforth referred to as White SAfE (WSAfE). This sociolectal continuum is traditionally broken up into three units, referred to by Lass (1995: 93) as "the great trichotomy" (a feature shared with other Southern Hemisphere Englishes):

1. A standard with an external British reference: in terms of pronunciation this is near-RP in Wells' (1982: 297–301) sense and often approximates an older form of RP. This variety is hardly used among young speakers any longer (Lass 2002: 110). This is referred to in the literature as either Conservative or Cultivated SAfE (henceforth CWSAfE).
2. A more local standard has progressively become the most widely spoken sociolect of WSAfE; in terms of accent, lexicogrammar, and lexis, this standard is distinctive in relation to other varieties of English. It is either referred to as Respectable or General SAfE (henceforth GWSAfE). According to some commentators, such as Lanham and Macdonald (1979), GWSAfE is, very roughly speaking, NE absorbed into the Johannesburg mixing process and reanalyzed as a sociolect. In Lanham and Macdonald's (1979) time at least both CWSAfE and GWSAfE were associated with "rejection of South Africanism in favour of links with the wider Anglo-Saxon world, a low level of patriotism, and hostility towards Afrikaners" (Jeffery 1982: 254). We suspect, however, that in the intervening 30 or so years, and in the case of GWSAfE, these associations have largely dissipated, partly as a result of the spread of GWSAfE at the expense of the other sociolects and partly because of the ideological effects of the political change to a fully democratic society in 1994. Still, while Coetzee-van Rooy and van Rooy (2005) find that black participants in their attitude study revealed a slight preference for the most educated (but still distinctively black) accents, the GWSAfE speaker was regarded very highly too, and certainly more highly than the Broad SAfE speaker (see below).
3. A variety alternatively known as Extreme or Broad SAfE (henceforth BWSAfE): the indexicality of this variety is more than just working class, an observation that, we suspect, remains as valid today as it was in Lanham and Macdonald's (1979) time. As explained by Jeffery (1982: 253–255), BWSAfE is associated with attributes such as being "tough, manly, sport-mad, sociable, patriotic and other things beside …. Ext SAE is loaded with political-ideological meaning as well as social: the South African tradition is to be not only tough etc. but also conservative, right-wing, authoritarian, unsympathetic to African aspirations …. Ext SAE speech reliably predicts such views … which are a significant part of the stereotype of the 'typical local man'. And indeed you do not have to be LC [Lower Class] to conform to the stereotype". It should also be noted that "the more extreme the variety is, the harder it becomes to distinguish it from second-language Afrikaans English" (Lass 2004: 373). For Lanham and Macdonald (1979) and other commentators, the idea is, very roughly again, that CpE was absorbed into the Johannesburg mix and reanalyzed as BWSAfE.

During the twentieth century this sociolectal continuum has dispersed geographically, largely doing away with the original regional lects (CpE and NE) and creating a typical Southern Hemisphere level of regional homogeneity. Generally, GWSAfE has spread at the expense of both BWSAfE and, in particular, CWSAfE.

While WSAfE was undergoing its formative process, English also spread to other communities in the country, giving rise to nonancestral varieties that are widely encountered in contemporary South Africa. These include South African Indian English (ISAfE), Cape Flats English/Colored English (C[f]E) and Black South African English (BSAfE). Of these varieties, ISAfE has become the native language of the vast majority of its speakers, while a substantial minority of the Colored community has also adopted English as its home language – 21% according to the official 2011 census (Statistics South Africa 2012).

English is the home language of 86% of South Africans of Indian ancestry according to the 2011 census (Statistics South Africa 2012). Indentured laborers were recruited from India to work on the sugar plantations of Natal in the second half of the nineteenth century, while a number of free Indians, mainly traders, also emigrated to South Africa during this period. A total of about 150 000 Indians moved to Natal between 1860 and 1911, and about half of them stayed in South Africa upon the expiry of the indentured contracts (Mesthrie 1995). These immigrants spoke a variety of Indian languages, both Dravidian and Indo-European, some features of which have determined the linguistic nature of ISAfE. English was introduced very gradually into the linguistic repertoire of these immigrants and their descendants, with limited education until the 1950s, alongside some informal contact beyond the classroom (Mesthrie 1992). After the introduction of general schooling, however, language shift was very quick: Mesthrie (1992: 31) notes that older siblings brought English home from the school playgrounds, enabling younger siblings to enter school with a fair command of English. In the period from the early 1950s to the 1970s, English became the first language of virtually the entire school-going population in the Indian community (Mesthrie 2002b: 340), as first documented by Bughwan (1970), who found that 90% of her 547 respondents claimed English as their strongest language.

English came to share a place with Afrikaans in the linguistic repertoire of the Colored people of the Western Cape, where "Colored" refers mainly to descendants of slaves (who were emancipated by the British in the first half of the nineteenth century), as well as children of inter-racial marriages and some descendants of the Khoi who lived in the area prior to the arrival of Europeans. During the course of the nineteenth century, widespread Afrikaans/English bilingualism developed in this community, although a distinctive variety of Afrikaans remained the dominant language for most. In the latter part of the twentieth century, and even more so in the early years of the twenty-first century, however, English has become the dominant language of individuals entering the middle classes (Malan 1996; McCormick 2002). In practice, while differences can be observed in the English pronunciation of English and Afrikaans native speakers respectively, there is a shared core of pronunciation features, some of which can be related to the Cape Vernacular Afrikaans dialect spoken in the same community (Finn 2004).

BSAfE is the most widely used form of English in contemporary South Africa. The roots of the variety can be traced back to nineteenth century mission education (Beck 1997; Hodgson 1997; Shepherd 1941). Mission education provided excellent opportunities to acquire native-like competence by the end of the nineteenth

century and continuing into the first half of the twentieth century (see, for example, De Klerk 1999), and was responsible for almost all education among black South Africans until the 1950s (Elphick 1997: 1). However, Hirson (1981: 220) notes that by the beginning of the twentieth century, the total enrolment of Africans in mission schools was still very small. It gradually grew to about 45% by the middle of the twentieth century (Booyse 2011b: 245), but the majority of children did not proceed beyond the second school year, while poor resources and overcrowding were the order of the day (Booyse 2011a: 202–205).

At this point in history, however, the situation changed dramatically: the new government, the National Party, implemented the Bantu Education Act. The government took control of all African schools (Hartshorn 1992; Hirson 1981; Booyse 2011b). This had two effects: "Under the new regime more children were accepted into schools, but the education was even inferior to that provided by the independent schools" (Hirson 1981: 227). This created a situation in (racially segregated) black schools where English was taught by Bantu-language speakers who themselves had limited training and command of English, and resulted in low levels of achievement, reinforcing features of home-language transfer (Lanham 1966). Just as the situation started to stabilize by the 1970s (as evidenced by the improved performance in school examinations), the final phase of political protest against the apartheid government was kicked off by the protest action in Soweto in June 1976. From this point onwards, education in townships reserved by law for black South Africans never quite returned to stability until the political transition in the early 1990s (Booyse 2011b: 257–262).

The cumulative effect of the twentieth century educational and political history of South Africa on the development of BSAfE is that a small elite close to the variety spoken by native speakers was removed from society, and a much more numerous group of relatively poorly educated speakers, with limited contacts beyond their own communities, developed in the second half of the century. However, English remained an important asset to the black community and it continued to be used in a range of functions (De Klerk 1999). Renewed claims about ownership of English started to emerge in the wake of the 1976 protest action in Soweto. This was articulated forcefully by public figures such as Mphahlele (1985) and Ndebele (1987) at addresses to the English Academy of South Africa. Since the political transformation of 1994, English has only increased in importance in the black community, while access to the language has also increased. Hence, in the present generation, significant changes are likely to occur.

The pronunciation features of South African English

White SAfE pronunciation has a number of distinctive characteristics. The area that has attracted the most attention is its vowels. A few consonantal properties have been identified, but nothing unique has so far been recorded in the literature as far as its suprasegmental features are concerned.

The following vowel features have been identified in WSAfE:

- WSAfE displays what has been commonly (and egregiously) referred to as the KIN-PIN Split by Wells (1982: 612–613). As shown in Bekker (2009), this is not a phonemic split at all but rather the entrenchment of allophonic variation in the KIT vowel. Basically in certain restricted contexts (e.g., after /h/) KIT is pronounced [ɪ], before tautosyllabic /l/ it is [ɤ], while in all other contexts it is [ə].
- Unlike Australian English and New Zealand English, WSAfE does not have a diphthongized FLEECE vowel (i.e., [əɪ] or thereabouts); even in BWSAfE it is a categorically monophthongal [iː] – a possible influence from Afrikaans.
- WSAfE does not participate as fully in the Diphthong-Shift and MOUTH-PRICE Crossover as do the other two Southern Hemisphere varieties (Wells 1982); i.e., at least in GWSAfE, MOUTH often has a similar starting point to PRICE (i.e., [ɐʊ] and [ɐɪ] respectively), FACE has a narrow diphthong (i.e., [eɪ]), while GOAT in GWSAfE is often fronted as opposed to lowered (i.e., [øʉ]). There is also much evidence of monophthongization in GWSAfE: FACE, as mentioned, is often narrow, GOAT is often subject to glide-weakening and PRICE is in fact considerably fronted and monophthongized in certain prestige varieties within GWSAfE (i.e., [praːs] for *price*). This, however, only underlines the notion that a PRICE-MOUTH Crossover is not a particularly prominent feature of SAfE. It is only in the broader idiolects that one finds a relatively fronted MOUTH onset (i.e., [æʊ]), backed PRICE onset (i.e., [ɒɪ] or monophthongal [ɒː]), and lowered onsets for FACE and GOAT (i.e., [ɐɪ] and [ɐʊ] respectively).
- WSAfE is often recognizable in terms of its substantially backed BATH vowel, which in the broader lects also shows lip-rounding (i.e., [ɑː] or [ɒː]); SAfE differs from Australian English and New Zealand English in this respect, both of which have a fronted BATH vowel, i.e., [aː]. Bekker (2012) makes a direct link between this feature of WSAfE and the importance of Johannesburg in the formation of SAfE.

Few unique consonant features have been identified, but the following is known:

- WSAfE displays allophonic variation between a clear and dark /l/, but there is no evidence of /l/-vocalization in the coda position (i.e., [jeɬ] not Cockney-like [jeʊ] for *yell*), and also Yod-Assimilation (e.g., [ʧʉːn] not RP-like [tjuːn] for *tune*). According to Bowerman (2004: 935), aspiration is not consistently present in voiceless plosives in syllable onsets.
- Broad WSAfE often displays features that can be linked to early Afrikaans influence (via CpE), e.g., obstruent (tapped) /r/ (e.g., [reːliː] for *really*), semi-rhoticity, and epenthetic schwa (e.g., [fələm] for *film*). The L2 English variety spoken by Afrikaans speakers (i.e., Afrikaans English) also shows evidence of syllable-final devoicing ([dɒk] for *dog*), although some of the contrast is retained by lengthening the previous vowel (van Rooy and Wissing 1996).

South African Indian English displays a variety of dialect-specific phonetic features, many of which are traceable to the original Indian substrate languages. However, Mesthrie (2004) simultaneously observes that many of the phonetic variants are similar to older (Cultivated) WSAfE values, which may suggest something about the early- to mid-twentieth century when much of the input to ISAfE was transferred from white native speakers, with subsequent isolation due to apartheid legislation.

- Indian SAfE shares the allophonic variation associated with the KIN-PIN "Split" in WSAfE, but in general shows less evidence of glide-loss. Characteristic vowel features include an unrounded RP-like NURSE vowel (i.e., [ɜː], different in this respect to WSAfE, which has [øː]), a GOOSE vowel that tends to be more back than in WSAfE, and a short diphthong in GOAT (in the region of [ʊo] rather than [ʊɐ] or [ʊə], as found in BWSAfE and GWSAfE respectively) (Mesthrie 2004: 956–959). The backer values for GOOSE have been retained by younger speakers even after the advent of a more integrated society (Mesthrie 2010).
- Consonantal features include occasional retroflexion of /t, d, n/, the realization of /f, v/ as [ʋ̥, ʋ] and /θ,ð/ as [t̪,d̪] (i.e., [d̪] for *then*), and unaspirated voiceless plosives in some environments (Mesthrie 2004: 959–962).

Cape Flats English is likewise characterized by certain conservative values, but shares the KIN-PIN "Split" with WSAfE and ISAfE (Finn 2004):

- Raised vowels, front and back, are characteristic of CfE, becoming more extensive as one moves further away from the prestige variety along the dialect continuum. Woods (1987) observes this for front KIT, DRESS, and TRAP as well as back LOT and THOUGHT. By contrast, he observes that STRUT is lowered to [a]. These features are also characteristic of all dialects of Afrikaans in the Western Cape. Wood (1987) also notes that unstressed vowels are not consistently reduced, but are often realized as peripheral. Finn (2004) points to the prevalence of "Canadian Raising" of PRICE and MOUTH with non-low onsets (i.e., [əɪ] and [əʊ]) in pre-fortis environments.
- Consonant features include an antedental /f/ (lower lip advanced beyond the top teeth), final-nasal elision ([plæ̃] for *plan*), and /h/ as voiced, i.e., [ɦ], the influence being conceivably both of a historical nature (in terms of language contact) and synchronic (in terms of L1 interference in the case of Cape Vernacular Afrikaans speakers (Finn 2004)).

Black South African English has been studied more extensively than any other nonancestral variety of SAfE. The general picture is one of a number of distinctive vowel and suprasegmental features, attributable to transfer from the native languages. Differences in consonants are fewer, and mainly due to phonotactic and syllabification differences. However, in recent years, there are clear indications of a gradual change in the pronunciation of some BSAfE speakers, in particular the so-called acrolectal group, which regularly interacts with native speakers and

other acrolectal speakers of nonancestral varieties. Nevertheless, there is as yet no overriding evidence of large-scale homogenization of upper class Black speakers and native (White, Colored, or post-acrolectal Indian) speakers.

- The picture that emerges from older descriptions of BSAfE vowel contrasts (e.g., Hundleby 1964; Lanham 1966; Adendorff and Savini-Beck 1993) is that the contrast between tense and lax (alternatively long and short) vowels is neutralized, and central vowels tend to be replaced by their closest front vowel alternative. Typical consequences of such mergers include the homophony of pairs like *sit* and *seat* (no length contrast) or *bird* and *bed* (central vowel replaced by front vowel). This is attributed to the constraint imposed by the Southern Bantu languages, which have five- or seven-vowel systems, but no phonemic length or tense/lax contrast. Van Rooy and van Huyssteen (2000), analyzing a small number of speakers acoustically, still largely confirm this picture, at least as far as monophthongs are concerned.
- Due to the relative absence of central vowels, there is no allophonic WSAfE-like KIN-PIN 'Split', although in the results of Van Rooy and van Huyssteen (2000) mid-front [ɛ]-realizations for the vowels from the PIN-set occur more frequently than for the KIN-set (even if [i]-realizations remain most frequent for both sets). This finding can be taken to suggest emerging awareness of the allophonic variation in the speech of BSAfE, without yet translating into a consistent articulatory replication.
- There is less agreement as to the realization of diphthongs in the older literature. While some sources simply claim the absence of diphthongs, Hundleby (1964) and Lanham (1966) in particular reported the breaking of diphthongs in bisyllabic sequences through glide insertion, resulting in PRICE being realized as [ajɪ] or MOUTH as [awu]. Van Rooy and van Huyssteen (2000) use acoustic data to show that in CHOICE the diphthongal realization is general, and some gliding is found in MOUTH, but that with the other traditional English diphthongs there is insufficient evidence for anything other than monopthongal realizations. However, drawing on a larger dataset of similar speakers, van Rooy (2004) adds PRICE and GOAT as potential diphthongs in BSAfE.

Research since the early to mid-2000s points to a new group of BSAfE speakers that exists alongside the speakers of older forms of BSAfE. Following a widespread practice in research on New Englishes, these two varieties are termed acrolectal and mesolectal BSAfE respectively. Such research began to observe changes in the speech of black South Africans about a decade after the political transformation of the early 1990s. The relevant observations about new phonetic realizations of vowels are the following:

- Starting with van Rooy (2004), researchers have observed the presence of lax vowels in the phonetic output of acrolect speakers, although at the time van Rooy studied these (drawing on data from 2000 to 2003), the lax vowels were

observed for both traditionally tense and lax monophthongs, i.e., FLEECE and KIT both showed the realizations [i] and [ɪ]. Da Silva (2007) also observes the emergence of central vowels in the speech of a subset of BSAfE speakers, especially for the KIT and NURSE vowels. Most recently, Mesthrie (2010), studying speakers at the very top of the socioeconomic spectrum who have been to fully integrated multiracial schools, reports that Black speakers in his sample approximate the WSAfE speakers' fronted GOOSE vowel to a stronger degree than Colored or Indian speakers of the same socioeconomic status. Like van Rooy (2004), however, Da Silva (2007) still observes a large group of speakers who produce vowels such as [ɛ] for NURSE or [a] for STRUT, thus not showing evidence of a tense/lax contrast.

- Diphthongs seem to have undergone a more extensive change in the speech of BSAfE speakers. Apart from the CHOICE diphthong already reported by van Rooy and van Huyssteen (2000), Da Silva (2007) reports that almost all of her BSAfE speakers share diphthongal realizations of MOUTH and PRICE, while many also have diphthongal realizations in GOAT and FACE.

Like the other varieties of SAfE, there are fewer unique consonantal features in BSAfE. The following features have however been reported in the literature:

- Consonant cluster simplification is attested in complex codas, especially when the syllable has a final plosive following another obstruent. If the following syllable starts with the same or a similar obstruent, the deletion of a coda obstruent is almost categorical. Resyllabification of coda plosives occurs in just less than 50% of the cases where the following syllable has no consonantal onset, but considerably less frequently in acrolect speakers. To a lesser extent, and only in the speech of mesolect speakers, the sonorant /r/ in onset clusters is deleted occasionally (van Rooy 2007).
- Final devoicing, without compensatory lengthening of the preceding vowel, is quite widespread (van Rooy and Wissing 1996, 2001).
- Like most other varieties of SAfE, BSAfE is nonrhotic. It also shows extensive aspiration of voiceless plosives, including in syllable codas (van Rooy 2000).

Suprasegmental features have also received some attention and are often the main target of prescriptivist comment on BSAfE. However, available data pertains only to mesolect speakers:

- BSAfE displays syllable-timed rhythm rather than stress-timed rhythm (Coetzee and Wissing 2007). In consequence, vowel reduction is not particularly prominent (van Rooy 2004). There is some debate as to what the typical realization is of vowels that are otherwise unstressed in WSAfE. Van Rooy and van Huyssteen (2000) find that in many cases the pronunciation is the mid-front vowel [ɛ/e], while low vowels such as [a/ɑ] are also common in final syllables, especially when the final syllable is open. Mesthrie (2005) argues for a more complex system of realizations.

- Stress patterns are different in mesolectal speakers. While there is a strong tendency to stress the penult, as noted by Lanham (1984), van Rooy (2002) finds that a super-heavy final syllable (or even a heavy final syllable) attracts stress to the final syllable, e.g., "realize" with final stress.

Recent developments and research into SAfE

There is growing evidence to suggest that SAfE might be undergoing a process of nascent regionalization, i.e., that speakers in the different English-speaking urban centers of South Africa (Cape Town, Port Elizabeth, Kimberley, Durban, and Johannesburg) are developing their own manner of speaker and indexing regional provenance. This appears to be true both of WSAfE (Bekker 2007; Bekker and Eley 2007; O'Grady and Bekker 2011) and other varieties such as ISAfE and C(f)E (Mesthrie 2010).

Of perhaps greater interest, however, are the linguistic reflexes of the growing racial integration that has taken place since the advent of full democracy in 1994 in South Africa. What integration exists has been mainly the result of a burgeoning black middle class, so it is particularly at this level of the social class continuum that new developments in SAfE have been noted. Van Rooy (2004, 2007) already identifies the presence of new variants in the speech of acrolect speakers, many of which are closer to WSAfE than the variants attested to in older/mesolectal BSAfE. Da Silva (2007), following Horvath (1985), uses a Principal Components Analysis to analyze the accents of students at the University of the Witwatersrand in Johannesburg and provides evidence for various changes within the English used by black individuals. More recently, Hartmann and Zerbian (2009) have shown that while middle class (particularly female) black South Africans often approximate GWSAfE, they are also, it would appear, creating new means for indexing ethnic identity; in this particular case Hartmann and Zerbian (2009) found evidence for neo-rhoticity (GWSAfE being a nonrhotic variety) in the speech of many such subjects. Research currently underway is investigating whether or not young white female South Africans are attempting to emulate their black peers in this regard. Mesthrie (2010) has broadened the investigation to include all ethnic groups (white, black, colored, and Indian) and concludes, in his study of GOOSE-Fronting among young middle class South Africans and with a few "ifs and buts", that "middle-class, L1 English-speaking South African students of all backgrounds are fronting the GOOSE vowel"; this is a sign of the possible development of a new, deracialized, middle class variety of SAfE.

At the same time, however, there are a number of similar features across the nonancestral varieties, ISAfE, C(f)E, and BSAfE, that may, with mutual reinforcement, remain resistant to convergence with the GWSAfE variety. The presence of a syllable-timed rhythm is reported for both BSAfE and ISAfE. Stress shifts to the right edge of the word are reported for BSAfE (van Rooy 2002), ISAfE (Mesthrie 2004a) and C(f)E (Finn 2004), with relevant examples being *realíze*, *intoxicáted*, and

participáte. While these authors use different terms, the actual examples they provide show how similar the process is across all three varieties.

Conclusion

South African English was transmitted to South Africa early in the nineteenth century by settlers from predominantly the south-east of England. While koineization was frequently interrupted by new waves of settlement, a stable form of SAfE must have been in place by the first half of the twentieth century. English spread very gradually and slowly to other communities, with elite bilingualism being a very noticeable part of early spread. However, after the introduction of general education by the apartheid government in the early 1950s, two major types of changes took place: the spread of English to other communities was accelerated considerably, even to the point of becoming the home language for the vast majority of the Indian community, but due to the isolation apartheid enforced between communities, distinct ethnolects developed. Only since the early 1990s have the boundaries that kept groups apart been removed, although the majority of especially the black community still lives in segregated areas, with limited contact with other speech communities. For those individuals of the South African community who are in the middle class or otherwise have access to more integrated educational facilities and an integrated workplace, there are early signs that within the first two decades of an open society, some degree of convergence between the various accents can be detected.

REFERENCES

Adendorff, R. and Savini-Beck, M. 1993. The teaching of English vowels and consonants in the new South Africa. *Journal of Language Teaching* 27: 232–248.

Beck, R.B. 1997. Monarchs and missionaries among the Tswana and Sotho. In: *Christianity in South Africa: A Political, Social and Cultural History*, R. Elphick and R. Davenport (eds.), 107–120, Cape Town: David Phillip and Oxford: James Currey.

Bekker, I. 2007. Fronted /s/ in General White South African English. *Language Matters* 38(1): 46–74.

Bekker, I. 2009. *The vowels of South African English*. PhD dissertation, North-West University. Available from: http://dspace.nwu.ac.za/handle/10394/2003.

Bekker, I. 2012. South African English as a *late* 19-century extraterritorial variety. *English World-Wide* 33(2): 127–146.

Bekker, I. and Eley, G. 2007. An acoustic analysis of White South African English (WSAfE) monophthongs. *Southern African Linguistics and Applied Language Studies* 25(1): 107–114.

Booyse, J.J. 2011a. The provision of education during the first half of the 20th century. In: *A History of Schooling in South Africa: Method and Context*, J.J. Booyse, C.S. le Roux, J. Seroto, and C.C. Wolhuter (eds.), 169–214, Pretoria: Van Schaik.

Booyse, J.J. 2011b. Education provision during the period of National Party rule. In: *A History of Schooling in South Africa: Method and Context*, J.J. Booyse, C.S. le Roux, J. Seroto, and C.C. Wolhuter (eds.), 215–268, Pretoria: Van Schaik.

Bowerman, S. 2004. White South African English: phonology. In: *A Handbook of Varieties of English: vol. 1: Phonology*, E.W. Schneider, K. Burridge, B. Kortmann, R. Mesthrie, and C. Upton (eds.), 931–942, Berlin and New York: Mouton de Gruyter.

Branford, W. 1994. English in South Africa. In: *The Cambridge History of the English Language, vol. V: English in Britain and Overseas: Origins and Development*, R. Burchfield (ed.), 430–496, Cambridge: Cambridge University Press.

Branford, W. 1996. English in South African society: a preliminary overview. In: *Focus on South Africa*, V. de Klerk (ed.), 35–51, Amsterdam: Benjamins.

Bughwan, D. 1970. An investigation into the use of English by Indians in South Africa with special reference to Natal. PhD dissertation, University of South Africa.

Coetzee, A.W. and Wissing, D. 2007. Global and local durational properties in three varieties of South African English. *Linguistic Review* 24: 263–289.

Coetzee-van Rooy, S. and van Rooy, B. 2005. Labels, comprehensibility and status in South African English. *World Englishes* 24(1): 1–19.

Conradie, C.J. 1986. *Taalgeskiedenis* [Language History], Pretoria: Academica.

Da Silva, A.B. 2007. South African English: a sociolinguistic investigation of an emerging variety. PhD dissertation, University of the Witwatersrand. Available at: http://wiredspace.wits.ac.za/handle/10539/4955.

De Klerk, V. 1999. Black South African English: where to from here? *World Englishes* 18: 311–324.

Elphick, R. 1997. Introduction: Christianity in South African history. In: *Christianity in South Africa: A Political, Social and Cultural History*, R. Elphick and R. Davenport (eds.), 1–15, Cape Town: David Phillip and Oxford: James Currey.

Finn, P. 2004. Cape Flats English: phonology. In: *A Handbook of Varieties of English: vol. 1: Phonology*, E.W. Schneider, K. Burridge, B. Kortmann, R. Mesthrie, and C. Upton (eds.), 964–984, Berlin and New York, Mouton de Gruyter.

Giliomee, H. and Mbenga, B. 2007. *New History of South Africa*, Cape Town: Tafelberg.

Hartmann, D. and Zerbian, S. 2009. Rhoticity in Black South African English. *Southern African Linguistics and Applied Language Studies* 27(2): 135–148.

Hartshorne, K. 1992. *Crises and Challenge: Black Education 1910–1990*, Cape Town: Oxford University Press.

Hirson, B. 1981. Language in control and resistance in South Africa. *African Affairs* 80: 219–237.

Hodgson, J. 1997. A battle for sacred power: Christian beginnings among the Xhosa. In: *Christianity in South Africa: A Political, Social and Cultural History*, R. Elphick and R. Davenport (eds.), 68–88, Cape Town: David Phillip and Oxford: James Currey.

Horvath, BM. 1985. *Variation in Australian English: The Sociolects of Sydney*, Cambridge: Cambridge University Press.

Hundleby, C.E. 1964. Xhosa-English pronunciation in the South-East Cape. PhD dissertation, Rhodes University.

Jeffery, C. 1982. Review of *The Standard in South African English and its Social History (Varieties of English Around the World: General Series Volume 1)*, by L.W. Lanham and C.A. Macdonald. Heidelberg: Julius Groot Verlag, 1979. *Folia Linguistica Historica* 3(2): 251–263.

Jeffery, C. and van Rooy, B. 2004. Emphasiser *now* in colloquial South African English. *World Englishes* 23(2): 269–280.

Kaplan, M. and Robertson, M. 1991. Introduction. In: *Founders and Followers: Johannesburg Jewry, 1887–1915*, M. Kaplan and M. Robertson (eds.), 10–16. Cape Town: Vlaeberg.

Lanham, L.W. 1966. Teaching English in Bantu Primary Schools. Final Report on Research in Johannesburg Schools. Publication No. 4 of the English Academy of Southern Africa, Johannesburg: English Academy of Southern Africa.

Lanham, L.W. 1984. Stress and intonation and the intelligibility of South African Black English. *African Studies*, 43(2): 217–230.

Lanham, L.W. and Macdonald, C. (1979) *The Standard in South African English and its Social History*, Heidelberg: Julius Groos Verlag.

Lass, R. 1995. South African English. In: *Language and Social History: Studies in South African Sociolinguistics*, Rajend Mesthrie (ed.), 89–106, Cape Town: David Philip.

Lass, R. 2002. South African English. In *Language in South Africa*, R. Mesthrie (ed.), 104–126, Cambridge: Cambridge University Press.

Lass, R. 2004. South African English. In: *Legacies of Colonial English: Studies in Transported Dialect*, R. Hickey (ed.), 363–386, Cambridge: Cambridge University Press.

Lass, R. and Wright, S. 1986. Endogeny vs. contact: "Afrikaans influence" on South African English. *English World-Wide* 7(2): 201–223.

Malan, K. 1996. Cape Flats English. In: *Focus on South Africa*, V. de Klerk (ed.), 125–148, Amsterdam: Benjamins.

McCormick, K. 2002. *Language in Cape Town's District Six*, Oxford: Oxford University Press.

Mesthrie, R. 1992. *English in Language Shift: The History, Structure and Sociolinguistics of South African Indian English*, Cambridge: Cambridge University Press.

Mesthrie, R. 1995. Language change, survival, decline: Indian languages in South Africa. In: *Language and Social History: Studies in South African Sociolinguistics*, R. Mesthrie (ed.), 116–128, Cape Town: David Philip.

Mesthrie, R. 2002a. Endogeny versus contact revisited: aspectual *busy* in South African English. *Language Sciences* 24: 345–358.

Mesthrie, R. 2002b. From second language to first language: Indian South African English. In: *Language in South Africa*, R. Mesthrie (ed.), 339–355, Cambridge: Cambridge University Press.

Mesthrie, R. 2004. Indian South African English: Phonology. In: *A Handbook of Varieties of English: vol. 1: Phonology*, E.W. Schneider, K. Burridge, B. Kortmann, R. Mesthrie, and C. Upton (eds.), 953–963, Berlin and New York: Mouton de Gruyter.

Mesthrie, R. 2005. Putting back the horse before the cart: The "spelling form" fallacy in Second Language Acquisition studies, with special reference to the treatment of unstressed vowels in Black South African English. *English World-Wide* 26(2): 127–151.

Mesthrie, R. 2010. Socio-phonetics and social change: deracialisation of the GOOSE vowel in South African English. *Journal of Sociolinguistics* 14(1): 3–33.

Mphahelele, E. 1985. Prometheus in chains: the fate of English in South Africa. *English Academy Review* 2: 91–104.

Ndebele, N.S. 1987. The English language and social change in South Africa. *English Academy Review* 4: 1–16.

O'Grady, C. and Bekker, I. 2011. Dentalisation as a regional indicator in General South African English: an acoustic analysis of /z/, /d/ and /t/. *Southern African Linguistics and Applied Language Studies* 28(1): 77–88.

Roberge, P.T. 1994. On detecting a prior linguistic continuum in Cape Dutch. In: *Nuwe Perspektiewe op die Geskiedenis van Afrikaans* [New Perspectives on the History of Afrikaans], G. Olivier and A. Coetzee, 153–165, Johannesburg: Southern Books.

Schneider, E. 2007. *Postcolonial Englishes: Varieties Around the World*, Cambridge: Cambridge University Press.

Statistics South Africa. 2012. *Census 2011: Census in Brief*, Pretoria: Statistics South Africa, Report 03-01-41.

Trudgill, P. 2004. *New-Dialect Formation: The Inevitability of Colonial Englishes*, Edinburgh: Edinburgh University Press.

van Rooy, B. 2000. The consonants of Black South African English: current knowledge and future prospects. *South African Journal of Linguistics* Supplement 38: 15–33.

van Rooy, B. 2002. Stress placement in Tswana-English: the makings of a coherent system. *World Englishes* 21(1): 145–160.

van Rooy, B. 2004. Black South African English: phonology. In: *A Handbook of Varieties of English: vol. 1: Phonology*, E.W. Schneider, K. Burridge, B. Kortmann, R. Mesthrie, and C. Upton (eds.), 943–952. Berlin and New York, Mouton de Gruyter.

van Rooy, B. 2007. Consonant clusters and resyllabification in Black South African English. *Language Matters* 38(1): 26–45.

van Rooy, B. and van Huyssteen, G.B. 2000. The vowels of Black South African English: current knowledge and future prospects. *South African Journal of Linguistics Supplement* 38: 35–54.

van Rooy, B. and Wissing, D. 1996. Degrees of neutralization during syllable-final devoicing: evidence from second language phonetics. *South African Journal of Linguistics* Supplement 33: 77–98.

van Rooy, B. and Wissing, D. 2001. Distinctive [voice] implies regressive voicing assimilation. In: *Studies on Distinctive Feature Theory*, T.A. Hall (ed.), 295–334, Berlin: Mouton.

Wasserman, R. 2014. Modality on Trek: modality in South African English. PhD dissertation, North-West University. Available at: http://dspace.nwu.ac.za/handle/10394/13042.

Wells, J. 1982. *Accents of English*, Cambridge: Cambridge University Press.

Welsh, F. 1998. *A History of South Africa*, London: HarperCollins.

Wood, T.M. 1987. Perceptions of, and attitudes towards, varieties of English in the Cape Peninsula, with particular reference to the "Coloured Community". MA dissertation, Rhodes University.

17 Indian English Pronunciation

PRAMOD PANDEY

Introduction

Indian English (IndE) represents one of the most prominent new Englishes (Mesthrie and Bhatt 2011). For a majority of its 125 million speakers, it is a second language, learnt at school and through higher education. For a small number of more than 200 thousand, in particular, the Anglo-Indian population, it is the first language, according to the 2001 Census of India figures. The pronunciation of English of Indians varies according to educational medium, level, and region, so that one can evidently speak of its variants such as Hindi English or Tamil English. English medium education as well as higher education has helped reduce the variation to the extent that a more general variety has emerged as an acceptable standard across the subcontinent, which has been given the name General Indian English (GIE).

My goal in this chapter is to illustrate the significant features of GIE and some of its variants. I first address the circumstances in which GIE has emerged as the representative variety of IndE. I then discuss the main features of the segmental and prosodic phonology of Indian English. I end the chapter with a brief discussion of an overview of issues relating to the stability of IndE.

English in India: past and present

GIE was proposed (CIEFL 1972; Bansal and Harrison 1974) as an educational standard for teaching English in India in place of British Received Pronunciation (RP), keeping in view the need to communicate with least interference of the mother-tongue with fellow Indians and foreigners. There has been a clear development from the stage of the introduction of English in India as a transplanted variety by the British rulers as a subject and medium of instruction in schools and colleges to its present stage as an educationally and socially accepted

The Handbook of English Pronunciation, First Edition. Edited by Marnie Reed and John M. Levis.

standard. This development, however, has not been without a competition with other varieties in the past, three of which were the most prominent, namely, Standard English, Anglo-Indian English (AIE), as used by the Anglo-Indian community of the pre-independence period, and a common variety popularly known as "Babu English" and "Butler English" (BIE for short) (Yule and Burnell 1996 [1886]; Hosali 2000) in its versions in Bengal and Madras presidencies respectively.

An example of AIE is provided in Allardyce (1877: 541) with "translation":

> Im dikk'd to death! The khansamah had got chhutti, and the whole bungla is ulta-pulta.
> [I'm bothered to death! The butler has got leave, and the whole house is turned upside down.]

An example of Butler English is given below from Hosali and Aitchison (2006: 57), the context is that of a butler on being invited to England:

> One master call for come India … eh England. I say not coming. That master very liking me. I not come. That is like for India – that hot and cold. That England for very cold.

As compared to the contact varieties of IndE illustrated above, there was a more standard variety, the "Indian English" in the early stages. The variety met with two opposite views: the purist and the realist views. The purist view aimed at bringing IndE close to the British standard, as expressed by Whitworth (1907: 5–6):

> I have been struck with the wonderful command which Indians – and not only those who have been to England – have obtained over the English language for all practical purposes. At the same time, I have often felt that what a pity it is that men exhibiting this splendid facility should now and then mar their compositions by little errors of idiom which jar upon the ear of the native Englishman.

The realist view, which dispassionately looks at the English produced by Indians with its own linguistic features and acceptable as such, must have existed for a long time, as is apparent from the title of Subba Rao's (1954) book, but it came to find expression in post-independence India. What has come to be widely accepted today as a representative variety of IndE is neither the Anglo-Indian variety nor the purist variety, but a variety that can function as a pedagogic model for acquisition through formal education. It is also assumed to be socially acceptable "devoid of regional peculiarities that may impair communication with speakers within and from outside the country" (Pandey 1994: 198). The rise of such a variety was possible because of a slow decline in the prestige of British Received Pronunciation (RP) "as a socially acceptable spoken variety of native English (nE), and a concomitant realization that it is too ideal a model for Indian learners of English to acquire". The wide acceptance of this variety throughout the subcontinent has been of such magnitude that the other contact variants are

seen today as aligned to it on a scale of acceptability. It is interesting to observe that Anglo-Indian English has come to converge with GIE, as noted by Coelho (1997) for one of the variants.

Elements of General Indian English Pronunciation: segments

In spite of some regional variation in its pronunciation, there is considerable stability in GIE to bear a description. We try to present such a description below based on existing studies.

We first present the segmental system, as described and discussed in CIEFL (1972) and later work.

Consonant phonemes

The consonant inventory is presented in Table 17.1 with some modifications from the inventories in CIEFL (1972) and Bansal and Harrison (1974).

In Table 17.1, the following points about the segment inventory of GIE may be noted. The labio-dental approximant /ʋ/ is substituted for two native English phonemes, the labial approximant /w/ (see also Sahgal and Agnihotri 1988) and the labio-dental fricative /v/, both of which are distinguished in restricted environments allophonically. The voiced alveolar fricative /z/ is substituted for both /z/ and /ʒ/, there being no post-alveolar voiced fricative in GIE. The following phonemes have different phonetic qualities from the segments in NE: /tʰ d ʈ ɖ ɾ / in place of /θ ð t d ɹ. /ɾ/ is variously termed as approximant or flap (Bansal 1976) or tap (as here). It has in fact variant pronunciations. The following phonemes with restricted occurrence have been added without corresponding

Table 17.1 Consonant phonemes of GIE.

	Bilabial	*Labio-dental*	*Dental*	*Alveolar*	*Post-alveolar*	*Retr.*	*Palatal*	*Velar*	*Glottal*
Plosive	p b (bʱ)		tʰ d			ʈ ɖ (ɖʱ)		k g (gʱ)	
Fricative		f		s z	ʃ				h
Nasal	m			n				ŋ	
Tap				ɾ					
Lat.				l					
Approx.									
Approx.		ʋ					j		
Affricate					ʧ ʤ				

Table 17.2 Monophthong vowel phonemes of GIE.

	Front	*Central*	*Back*
Close	ɪ iː		ʊ uː
Close-mid	e/ɛ eː	ə	oː
Open-mid	æ		ɔ (ɔː)
Open		aː	

consonants in NE: /bʱ/ as in *abhor*, /ɖʱ/ as in *adhere*, and /gʱ/ as in *ghost*. These consonants are based on the orthographic representations of the words in which they occur.

The distribution of the consonant phonemes in general is in accordance with the standard international varieties, with the following peculiarities. / ɾ / optionally does not occur before consonants and word-finally. The optional deletion of /ɾ/ is restricted to the word domain. If dropped finally, it does not surface, even when the following word begins with a vowel, e.g., /də ɾivə iz floːiŋ/ *The river is flowing*. Word-medially, /ŋ/ always occurs with a following /g/, e.g., /siŋ/ *sing*, but /siŋgiŋ/ *singing*. Geminate consonants occur within morphemes, e.g., *innate* /inneːt̪/, *happy* /hæppiː/, as well as across morphemes, e.g., *illegal* /illiːgəl/ *unnatural* /ˈənnæt̪ʃuɾəl], when the orthographic word has corresponding double consonants. The occurrence of geminate consonants in this context is, however, not consistent; it does not occur, for example, in *attest*, *rabbit*, *added*, etc.

The distribution of some of the phonemes in morphophonemic alternations (e.g., Bansal 1983) is at variance with native English, in a majority of cases, on account of orthography. The regular past tense allomorph is pronounced /ɖ/ or /ɛɖ/, as in *walked* /wɔːkɖ/, *robbed* /rɔːbɖ/, laughed /laːfɖ/, *wanted* /ʋaːnt̪ɛɖ/. The regular plural morpheme is pronounced /s/ for both /s/ and /z/, and /ɛz/in place of /ɪz/ of native English, e.g., *dogs* /dɔːgs/, *falls* /fɔːls/, *matches* /mæʧɛz/.

Consonant allophones

Aspirated plosives and affricate [pʰ t̪ʰ kʰ ʤʰ] < /p t̪ k ʧ/ are occasionally heard in the speech of educated Indians. The retroflex nasal [ɳ] < /n/ occurs before retroflex stops, e.g., [pæ̃ɳt̪s] *pants.* The retroflex stops interestingly have alveolar allophones following alveolar fricatives /s z/, e.g., [best] *best*, [reːzd] *raised*. In the variety encountered in the south the aspirated dental plosive [t̪ʰ] (for θ) is often unaspirated (e.g., Nagarajan 1985; Indira 2009) and the alveolar lateral tends to be retroflexed [ɭ] intervocalically (e.g., Indira 2009). Apart from the specific allophones, a general allophonic process of gemination found across Indian languages (Hindi, Bengali, Punjabi, Marathi, Tamil, Telugu, etc.) leads to geminate consonants in the environment between a vowel and /j ʋ r l/, e.g., *between* [biɪt̪wiːn], *supreme*

[suppri:m], *secure* [sekkjo: ɾ]. These cases of geminates are in addition to the ones on account of the double consonants in the orthographic forms of words, as observed above.

Vowel phonemes

The inventory of monophthong vowel phonemes is presented in Table 17.2.

Wells (1982) illustrates the differences among these vowels in terms of a lexical set – KIT, FLEECE, DRESS, FACE, and TRAP for the front monophthongs, FOOT, GOOSE, GOAT, LOT/CLOTH for the back monophthongs, and STRUT, PALM for the central monophthongs.

The diphthongs that occur on the surface are six in number: /aɪ aʊ ɔɪ ɪə ʊə/, occurring in the following lexical set of Wells (1982): PRICE, MOUTH, CHOICE, NEAR, SQUARE, CURE. For the last word, POOR would be a better example, as in the case of CURE, one can come across variants such as o: /kjo:(r)/. Although, on the surface, all the six diphthongs occur, three of them are more stable, namely, /ai au ɔɪ/.

The following facts of correspondence with a native variety such as RP, as described, for example, in Gimson (1962), may be noted. The monophthongs /e: o:/ correspond to NE diphthongs /ei əu/. The distinction between the long and the short low-mid back vowels /ɔː ɔ/ is neutralized to /ɔː/. A shorter variant /ɔ/ is also found, but does not contrast with the long counterpart. The realizations of the vowel are more restricted in IndE, giving way to the close-mid vowel /o:/ in certain words; e.g., *cot* /kɔ:ʈ/~ /kɔʈ/, *caught* /kɔ:ʈ /, *call* /kɔ:l/, *core* /ko:r/ (in place of RP /kɔ:(r)/), *court* /ko:(r)ʈ/, *coat* /ko:ʈ/. The vowel /o:/ has a wide distribution, occurring in place of the RP /ɔ:/, as mentioned above, as well as the diphthong /əʊ/: *go*, *court*, *road*, *force*, *more*, etc. /æ:/ is realized as [ɛ:] in most instances. Where full vowels alternate with schwa in native English varieties in stressed and unstressed positions, there is frequent occurrence of full vowels in GIE, even in those positions that are not stressed in words, such as in the underlined vowels in *acid/a̲cidity*, *photo̲graph/photogra̲phy*, *oppo̲se/oppo̲site*, *baseme̲nt*. Word-finally /a:/ occurs commonly in place of the NE /ə/, /pu:na:/, *Poona*, /ɪnɖɪa:/, *India*. Nonfinally, in nonalternating cases, /ə/ tends to occur frequently, e.g., *a̲bove*, *drive̲r*, etc. /ə/ is optionally deleted in unstressed syllables flanked by a preceding stressed syllable, and followed by another syllable, after a general Schwa Deletion process in Hindi (Pandey 1990), e.g., [miˈlitrɪ:] *military*, [sekˈkretrɪ:] *secretary*. In regional varieties of IndE, such as Bengali English and Tamil English, the nonalternating /ə/ may be realized as a front vowel /æ/ or /ɛ/, especially since the filter languages (Bengali and Tamil) have word-initial stress.

The distribution of the diphthongs on the surface is for the most part as in native English varieties, with occasional restrictions, such as (see Bansal 1983), /ɪə/ may be realized as a monophthong /i:/, /eə/ as /e:/ and /ʊə/ as /u:/ in certain lexical items, e.g., *serious* /si:riəs/, *period* /pi:riəɖ/, *area* /e:rɪa:/, *various* /ʋe:rɪəs/, *during* /ɖju:rɪŋ/, and *tour* /ʈu:r/. Although, on the surface, these six

diphthongs occur in GIE and in most varieties of IndE, it is difficult to establish that they are indeed single vowel phonemes and not a sequence of two vowels. One of the ways to ascertain their diphthongal status is by examining their behavior in some form, such as stress placement. In nE, for instance, the complex words *severity* /sɪˈverɪti/ and *theˈatrical* /θɪˈætrɪkl̩/ from *severe* and *theatre* provide evidence for the surface diphthongs in the shorter words to be a diphthong in one (*severe/severity*) but a vowel sequence in the other (*theatre/theatrical*). Pandey (1980) on the basis of the study of word-stress in Hindi English found that the placement of stress in words reveals that phonologically some of these diphthongs behave like vowel sequences, as is apparent from the following examples for the vowel /ɪə/: *severe* /ˈsi:ʊɪə(r)/, *severity* /sɪʊˈæriʈi:/, *sincere* /ˈsɪnsɪə(r)/, *sincerity* /sɪnsɪˈɛriʈi:/. Most major Indian languages from the Indo-Aryan and Dravidian stocks have limited number of diphthongs, /əɪ/ and /əʊ/ being the commonest. They differ, however, with regard to the presence of hiatus (i.e., occurrence of vowel sequence) in them. Hiatus is permissible in the Indo-Aryan languages, but is absent in some of the Dravidian languages, for example, Tamil. The occurrence of surface diphthongs in most varieties of IndE may in all likelihood be perceived as vowel sequences. It is relevant to recall here the tendency, as noted above, in IndE to substitute diphthongs with monophthongs in certain forms, such as *se̲rious*, *va̲rious*, *to̲ur*. In such cases the vowels must be assumed to be perceived as single and not a sequence.

A consideration of the inventory of diphthongs in GIE, with two of the RP diphthongs absent, namely, /eɪ/ and /əʊ/ or /oʊ/, draws one's attention to a historical common beginning of English in India in the seventeenth to eighteenth centuries. A look at the development of diphthongs in English, as discussed in detail in Dobson (1968), shows that the vowels /eɪ/ and /oʊ/ were the last to develop in the later eighteenth century. When English was transplanted in India, the diphthongs had not emerged. Any explanation other than the one based on historicity, such as universals or markedness, for the absence of the two diphthongs in most varieties of IndE can be as plausible.

Vowel allophones

The vocalic allophones of GIE differ to a much greater extent than the consonant allophones from other varieties of English in terms of their phonetic realization. Almost each vowel is different in quality from RP. Within the phonemic system of GIE, however, there is less allophonic variation. One of the main vocalic allophones is nasal vowels. Vowels are nasalized when they both precede or follow a nasal, e.g., [nõ:] *No*, [nɔ̃ʈ] not, [ẽnĩ] *any*. Nasalization does not take place if a voiced non-nasal consonant follows the vowel, e.g., [nɔ:d] *nod*, [mo:ɾ] *more*. Unlike in many varieties of native English, vowel length is insensitive to the voiced/voiceless distinction in the following environment. Thus /aɪ/ in *rice* and *rise* or *right* and *ride* is not differentiated in terms of length in its allophonic manifestations in the words.

Experimental evidence for GIE

Experimental phonetic studies of IndE sounds are of some standing now. However, a majority of them (e.g., Balasubramanian 1972, 1975; Gupta 1982; Nagarajan 1985; Indira 2009) are of regional varieties. Phonetic studies of a general nature are relatively recent. They are aimed at examining impressionistic as well as experimental observations of phoneticians regarding IndE sounds, in the main vowels and prosodic phenomena of stress and intonation. Wiltshire and Harnesberger (2006) investigate IndE pronunciation of two groups with different first languages, namely Gujarati, an Indo-Aryan language, and Tamil, a Dravidian language, "to evaluate to what extent Indian English (IE) accents are based on a single target phonological-phonetic system (i.e., General Indian English), and/or vary due to transfer from the native language." The investigation reveals "…both phonetic and phonological influences of IndE speakers' native language on their accent in IndE, even in proficient speakers; these influences appear to supersede IndE norms and can be found in both the segmental and suprasegmental properties of their speech."

Wiltshire and Harnesberger find that the observations of Wells (1982) and Nihalani et al. (1979/2004) regarding the low vowel /a/ being "front" in quality is not attested in the IndE of speakers with Tamil and Gujarati as their mother tongue. They use a back vowel /ɑ:/ instead. That result is further confirmed in Wiltshire (2009). The view that the phonemic status of the vowel /ɐ/ (earlier symbolized as /ʌ/), /ɜ:/, and /ə/ in IndE is not clear (e.g., Wells 1982), and that the former often appears even in unstressed syllables (e.g., Nihalani et al. 2004) is examined in Wiltshire (2009) and found to be attested. The vowel in fact occurs commonly in western Hindi and western Indo-Aryan languages such as Haryanvi and Punjabi.

The uncertainty regarding the pronunciation and structure of diphthongs in IndE noted above has confirmation from an acoustic investigation in Maxwell and Fletcher (2010), who base their studies on the L1 speakers of Hindi and of Punjabi. The data show that "none of the speakers produced a full set of diphthong vowels", with "mother-tongue interference as a relevant factor". The study, like most studies on diphthongs in IndE, does not discuss the issue of a distinction between diphthongs and sequence of vowels and focuses on surface pronunciation.

Prosodic features

Studies of IndE phonology generally acknowledge the significance of the prosodic phenomena in lending it its character. Bansal (1976) mentions wrong placement of stress/accent in words to be the most significant factor affecting the intelligibility of IndE to speakers of British English. Gumperz (1982) and Pickering (1999) show how intonation in IndE can be the source of misunderstandings at discourse level. However, studies of the prosody of IndE tend to be narrowly

focused on individual varieties, for example, Hindi English (Pandey 1980), Malayalee English (Nair 1996), Marathi English (Gokhale 1978), Tamil English (Vijaykrishnan 1978), and Telugu English (Babu 1974). Consequently, for an understanding of prosodic organization in IndE, a specific variety has to be taken as a case of instantiation of GIE.

Word-stress

The word-stress system in native English is significant on many counts. It has been shown to be clearly a lexical phonological phenomenon, interacting with morphology, and having many exceptions. Besides, the realizations of segments are often affected by the syllable being stressed or unstressed. These features are in general not found in IndE. I present below a detailed list of the patterns of word-stress in Hindi English. The asterisk indicates the difference from NE patterns. For ease of discussion, the words are presented in subgroups.

(1) **Verb**

e'merge	su'rrender	'diminish
e'lect	'differ	*'develop
a'dopt		*'solicit
di'vide		

Adjective

se'cure	*si'nister	*'terrific
di'vine	*se' mester	*'prolific

Noun

a'larm	u'tensil	'benefit
sa'loon	a'sylum	A'merica

(2) (i) compe't[iː]tion i'r[ɔː]nic
eco'nomic

(ii) *e'xaminee *'tattoo
*'degree *'shampoo
**'cassette *'canoe

(3) (i) *'defer *'deter
*en'gin[ɪə]r~ *'engi'neer *bio'l[ɔː]gy
*'career *'astro'nomy (~)
*'cashier *'sincere
*'[iː]vent ~ e'vent *'med[i]val
*em'ph[æ]sis *'comfor'table

(ii) *so'litary *se'cretary ~ 'secretary
*ca'tegory~ 'category

(iii) *ˈphotoˈgraph *ˈphotoˈgraphy
*ˈphotoˈgraphic *ˈphotoˈgrapher
*ˈpentaˈmeter *ˈhexaˈgonal
*ˈmetaˈthesis

(iv) *masˈterly *proˈperly

(v) *ˈmoreover *ˈhowever

(4) ˈpermit$_N$ *ˈpermit$_V$
*exˈport$_N$ exˈport$_V$

(5) *eˈxamiˈnation (~) *ˈinterroˈgate (~)
*aˈttesˈtation (~) *ˈacclimaˈtize (~)
*ˈciviˈlize (~) *ˈqualiˈfy (~)

(6) *ˈloud{speaker *ˈbad-ˌtempered
*ˈsecond-ˌclass *ˈthree-ˌwheeler

The following generalizations hold for word-stress in Hindi English (HE).

(7) a. Most stress patterns are phonologically predictable. Thus those stress patterns in NE that are phonologically not predictable but lexical, such as *caˈsette*, *caˈnoe*, *deˈgree*, etc., are regular in HE: *ˈ*cassette*, *ˈ*canoe*, *ˈ*degree*, etc.

b. There are many instances where the stress patterns appear to be lexical in HE, as in the following: ˈ*event*, ˈ*medieval*, etc. Their apparent lexicality in such instances is on account of **restructuring** in the underlying representations of the words, involving a long vowel being short or vice versa, e.g., ˈ*[i:]vent*,ˈ*med[i]val*, etc.

c. There are instances in which the stress patterns in HE are the same as in nE, but which involve a change in the phonemic status, e.g., *compeˈt[i:]tion*, *iˈr[ɔ:]nic*, etc.

d. Complex words, with two stresses, as in (5), do not have fixed primary stress, following a general pattern in Hindi (e.g., Pandey 1989). Either the first or the second stress may be primary.

e. The compound words, contrary to the general pattern in complex words, have a fixed pattern. The first member of the compound has primary stress. (A source of difficulty, also pointed out in Gopalakrishnan 2011, is that both compound stress and phrasal stress on Modifier + Head constructions alike have primary stress on the first word and secondary stress on the second word, thus the compound ˈ*white ˌhouse* and the phrase *(a)* ˈ*white ˌhouse* are pronounced alike.)

Phonetics of word-stress

Phonetic studies of word stress in IndE (e.g., Mohanan 1986; Pickering 1999' Pickering and Wiltshire 2000; Wiltshire and Moon 2003) agree on one main feature of the realization of stress, namely, fall in pitch on the stressed syllable. Wiltshire and Moon (2003) conducted a production and perception study on phonetic correlates of stress in Hindi and the differences between American English (AmE) and IndE. The speakers, 10 AmE and 20 IndE, were given 60 words to produce in a carrier sentence "I will say X again". The results show that in addition to a fall in F0, two other correlates, namely, increase in amplitude and duration, go with stressed syllables, but they are not significant. Following Beckman (1996), the authors term IndE as a "pitch accent" language, like Japanese, in which a fall in F0 is the main phonetic correlate, and amplitude and duration do not play any role.

Although the observation regarding a fall in F0 is attested even for the substrate Indic languages, such as Hindi (Dyrud 2001) and Tamil (Keane 2005), a recent study by Féry, Kenntner, and Pandey (2013) offers a different interpretation of the facts. As the latter study is on Focus, it is discussed in a later section below.

Rhythm and intonation

It has been generally held (e.g., Bansal 1969; Wells 1982) that IndE, like its substrates, has a different rhythm and intonation system than the native varieties of English. While this general assumption has been time and again found to be true, the exact nature of the rhythm and intonation of IndE as well as of Indic languages in general is in need of investigation.

The speech rhythm of IndE is generally labeled as "syllable-timed" (e.g., Kachru 1983; Gargesh 2004) compared to native English, which is "stress-timed". The senses in which the terms are used are that in stress-timed languages the duration between stresses is roughly equal, irrespective of the number of syllables in them, and that in syllable-timed languages, the number of syllables determines the duration of spoken units. As Adams (1979) points out, the duration of the stretches between ˈ*ma-* and ˈ*here* in the sentences *The*ˈ *manager is* ˈ *here* and *The*ˈ *man is* ˈ *here* are roughly equal in the speech of a native English speaker, and different in the speech of the speaker of English with a syllable-timed rhythm. While the sense in which the term "stress-timed" is applied to native English is held to be valid, the sense in which the term "syllable-timed" is applied to IndE is not valid on two grounds. One, the definition has been found to be controversial (Roach 1982) and to be inapposite for languages such as French (Wenk and Wioland 1982), Tamil (Balasubramanian 1980), and Telugu (Babu 1971). There has been an attempt at resuscitating the distinction by redefining the terms (e.g., Dauer 1983, 1987; Auer 1993; Schiering, Bickel, and Hildebrandt 2012). For lack of space, we cannot go into the renewed distinction here. It is relevant to note, however, that the general logic on which a distinction between languages in terms of rhythm (Ohala and Gilbert 1979) is needed finds support from studies on speech perception (Auer 1993).

One of the consequences of stress-timed rhythm is that in order to maintain consistency in the duration between stresses, unstressed syllables tend to be reduced, as can be seen in the related forms *photo*, *photograph*, and *photographer*. The underlined vowels are reduced when unstressed. The unstressed vowels are even deleted when not stressed, as in *we've*, *they're*, etc. Words such as *have* and *are*, known as Function Words, are in general not stressed; they are stressed in restricted contexts. They are thus known to have two forms, strong and weak. In IndE, related forms like *photo*, *photograph*, and *photographer* tend to have full vowels once one of them is stressed and function words tend to have vowels pronounced in full, giving the impression of pronouncing them as strong forms. The need for sufficient and regular pronunciation of weak forms has been expressed by many (Ladefoged 1993; Wells 2000; Roach 2001) in order to avoid miscommunication. Most studies on the varieties of IndE point to the general tendency for pronouncing the function words in their strong forms. Madhavi (2009) reports a reading test of 26 function words and 8 contracted forms pronounced in the positions for their pronunciation as weak forms conducted for 20 students (11 male and 9 female) in the age group 20–23 years pursuing MBA studies through the English medium. The results showed that out of the 34 forms, 10 were pronounced with their weak forms or contracted forms by 5% of subjects and only 2 by 10% of subjects. Of the function words and contracted forms 22 were pronounced in their strong forms by 100% of subjects; 10 by 95% of subjects, and 2 by 90% of subjects.

The intonational studies of IndE usually relate it to the substrate Indic mother tongue for explaining the patterns, and, indeed, there are similarities between them. Studies reported in Latha (1978) and Nair (2004) for Malayalee English, Wiltshire and Hanrserger (2006) for Tamil English and Gujarati English, Babu (1974) and Joseph (1984) for Telugu English, Gokhale (1980) for Marathi English, and Khan (1974) and Shekhar (1993) for All-Indo Radio announcers of IndE in general point to two common features – one, the presence of multiple stresses in an intonational unit and, two, the placement of the nucleus on the last but one word in an intonational unit with Modifier + Head constructions. Some examples for tonal placement patterns in the speech of All India Radio News readers are reproduced in (8) from Shekhar (1993: 50):

(8) 1. …in ˈlieu of the ˌquestion hour// (nE: … in ˈlieu of the ˈquestion ˌhour//)
2. … and ˈseek a ˌfresh mandate// (nE:… and ˈseek a ˈfresh ˌmandate//)
3. …in the forty-ˋeighth over// (nE:… in the ˈforty-ˈeighth ˌover/…)

The studies on intonation in the varieties of IndE show that while IndE intonation differs from native English intonation and in that sense has a unity and its own identity, there is internal variation among its varieties. One common feature is the occurrence of prominence on function words, as discussed above. The other important feature, as discussed by Wiltshire and Harnsberger (2006), is the occurrence of many more pitch contours assigned to words in an intonational phrase than is normal in native English. In an analysis of read sentences by speakers of Tamil English (TE) and Gujarati English (GE), they found that in both varieties all

content words were assigned a pitch accent, with the speakers of the two varieties using different pitch accents. "GE speakers typically use a rising pitch accent transcribed here as LH, while TE speakers use either a falling pitch accent (HL), a high pitch accent (H), or a rising pitch accent (LH)" (Wiltshire and Harnsberger 2006: 101). The assignment of rising patterns in GE is very similar to that observed by Rajendran and Yegnanaraya (1996).

Types of tones

We do not have an exhaustive account of types of tones used in IndE. However, two opposite cases claim our attention, namely, Gokhale (1978) and Latha (1978). One shows tonality to be fairly similar to RP and the other shows it to be fairly different from RP. Gokhale (1978, 1980) mentions the following tones used in Marathi English: three simple tones – Fall, Low-Rise, and High-Rise – and one complex tone – Fall-Rise. The senses in which they are used are fairly similar to the senses in Marathi, and broadly in RP. The High-Rise in Marathi English is especially found in echo questions and yes/no questions. The latter are also said with a Low-Rise. According to Gokhale (1978: 172), "… a speaker of Marathi English does not have much difficulty in acquiring the patterns of tonality in R.P".

Latha's (1978) description of Malayalee English posits the following tones: one simple tone – Fall – and four complex tones – Rise-Fall, Fall-Rise, Drop-Rise, and Drop-Rise-Fall. The tonality of Malayalee English is obviously quite different from the tonality of RP. Notice that a simple Rise tone is virtually missing in Malayalee English, in which a Rise begins with a drop.

The use of the nuclear tone in both varieties is stated to be on the last lexical word, except when the clause final NP has a Modifier + Head structure, in which case the tonic is on the penultimate word. This appears to be a general pattern in IndE, as pointed out above.

Information structure: focus

Féry, Pandy, and Kentner (2013) report the results of an experimental study conducted on Hindi and IndE speakers to investigate the prosodic correlates of focus, by eliciting data containing focused and given words. The data were elicited in the form of recordings of semi-spontaneous speech in response to a task of the QUIS questionnaire (Skopeteas et al. 2006), called "Anima". In the theoretical framework used, prominence and alignment were seen as two separate parameters of focus expressions. The focused elements in the data on IndE showed one or a combination of the following correlates of prominence: higher F0 on Object focus expressed as L*H melody on nonfinal elements and as H*L melody on final elements, a "hammock"-like structure expressed as a dip and a rise back to the level before the dip in the F0, giving an H*LH melody, an increase in amplitude, and an increase in duration. The last two were not significant. More

conspicuously, focus was found both in Hindi and IndE to be accompanied by stronger phrasal correlates. Both Hindi and IndE were found to align focus to the left of a phonological phrase (roughly seen as an intermediate phonological constituent between a phonological word and an intonational phrase or unit; see Selkirk 1984).

The following illustration from the data collected for the study (Féry, Pandey, and Kentner Pandey 2013) shows both the aspects of prosody of focus in IndE – prominence and alignment (see Figure 17.1). The intonational phrase (IP), *a girl is hitting a boy* with Object focus, shows the LH melody on the Subject, with L at the beginning of the phrase *a girl* and H at the end. The clause is in answer to the question, "In the garden, Is the girl hitting a girl or a boy?"

The IP *A girl is hitting a boy*, with Object focus, shows that the prosodic alignment on the phrases are different from NE, where the phonological phrases show different groupings among words. There are three phonological groups – *a girl is*, *hitting the*, and *boy*. The pronunciation of IndE is rendered different from NE with the alignment of the LH melodies on the first two groups and the separation of the third group "boy" with focus on it. Studies on other varieties of IndE (e.g., Latha 1978 for Malayalee English) show that the phrasing of prominence and focus is a characteristic feature of IndE with major Indian languages (e.g., Féry, Pandey, and Kentner 2013 for Hindi, Mahesh 2014 for Malayalam).

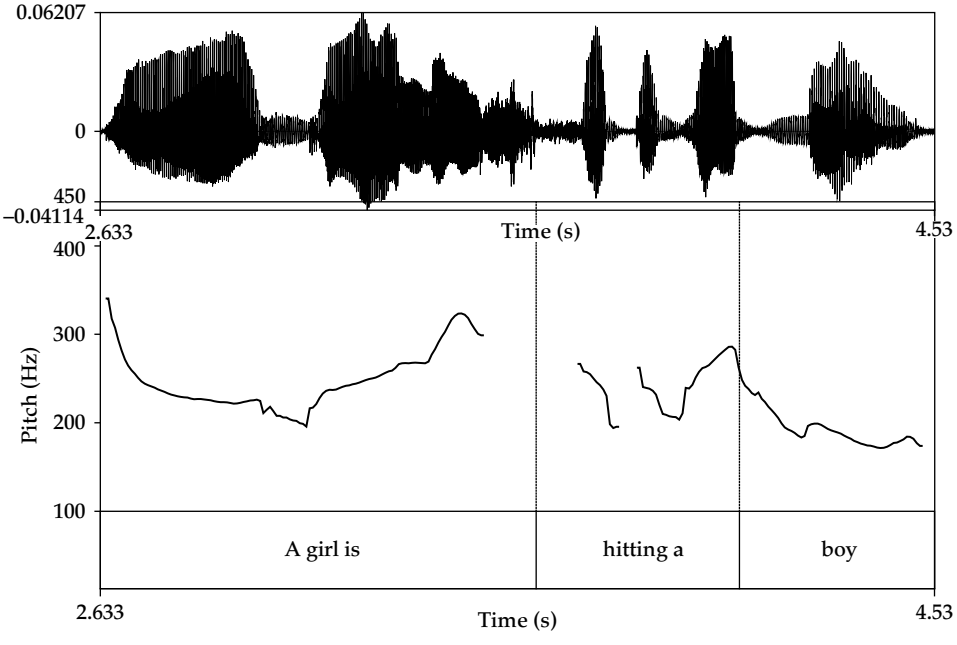


Figure 17.1 LΦHΦ HΦLι
(A girl is)Φ ((hitting a) boy)Φ)ι

Stability

Until recently, approximately around the mid 1970s, the general attitude towards IndE pronunciation was as a deviant variety. It is common to find expressions such as "faulty (location)", "wrong (placement)", "deviation", etc., in the studies of IndE pronunciation until that period. The trend is reversed today, especially from the time of the proposal of the idea of GIE within India and of IndE as an instance of non-native varieties (e.g., Kachru 1982). IndE pronunciation is seen today as a variant of English pronunciation. Along with other features of grammatical and discourse structure, it has even come to be used as a basis for questioning the distinction between "native" and "non-native" varieties of English (e.g., Agnihotri and Singh 2012).

English in India has been seen more as a medium of higher education than as a medium of mass literacy in India. It would appear that for this reason it has not been the focus of language policy in India. The attention of policy makers has been divided between the regional languages, the indigenous languages, and the official languages, Hindi and English. This could well explain the features of IndE as far from the character of a "high" language, even though it functions today on a par with Sanskrit in ancient times (see Dasgupta 1993).

The development of IndE is expected to take place as a natural system and a living force. Its institutionalization is already taking roots, based on a general assumption about many of the common features among its variants, at both segmental and prosodic levels. The common segmental phonological features include the following among those noted above – the presence of retroflex stops in place of alveolar stops (except in the north-eastern variety), dental plosives in place of dental fricatives, neutralization of the vowels back rounded mid and low vowels, and the mid monophthongs /eː oː/ in place of diphthongs /eɪ əʊ/. The following can be mentioned among the common prosodic features – the absence of lexical conditions in word-stress patterns and the predominance of phrasal units in an intonational unit (see, for example, Féry, Pandey, and Kentner 2013) and a greater tendency towards syllable-timing in speech. When we examine closely all these common features, which also function as the acceptable features of pronunciation for the speakers of the regional varieties of Indian English, strongly suggest GIE to be a contact variety (Pandey 2014).

There is evidence for its institutionalization beginning to take place in language technology research that already recognizes IndE pronunciation, e.g., Sen and Samudravijaya (2002), Sen (2003), Kumar, Kataria, and Sofat (2003), Mullick et al. (2004), and Kumar et al. (2007). Studies such as these in the field of automatic generation of IndE speech provide useful insights into the similarities and differences between native English and IndE pronunciation. Thus Kumar et al. (2007) show that automatic generation of Indian pronunciation of English words to the baseline Carnegie Melon University dictionary showed the need for only 26.3% of words needing correction against standard native English pronunciation. Of these, 19.1% differences were at the prosodic level (mainly word-stress) and only 7.2%

differences required phoneme substitutions for being usable as IndE Voice. Of these, the most common substitutions included vowels (/a:/ ➔ /ɔ/, e.g., *hostilities*), as well as consonant substitutions (like /z/ ➔ /s/ and /w/ ➔ /v/, as reported). The /v/ here may be /ʋ/ in all likelihood. The figures described here are for pronunciation lexicon at the word level. It should be obvious by now that the difference is expected to widen at the level of sentence prosody.

Conclusion

In the present chapter, we began with looking at the development of GIE as a well-considered pedagogic choice. The considerations, however, have been found to be explicit on the segmental aspect of pronunciation, but lacking in a definite form on the prosodic aspect. Following this discrepancy in its description, the segmental and the prosodic elements of pronunciation in GIE were presented in separate sections. The current trend towards the stability of IndE was taken up for a brief deliberation in the end. For lack of space it was not possible to delve into certain aspects of IndE pronunciation, such as regional variation in the realization of segments, given the wide variety of the substrata and the organization of the sound system in terms of markedness considerations. These are desiderata for future research.

Acknowledgment

The work reported here was supported by a traveling grant from the School of Language, Literature and Culture Studies of Jawaharlal Nehru University in June 2012, enabling the author to consult libraries for doctoral dissertations on Indian English pronunciation.

REFERENCES

Adams, C. 1979. *English Speech Rhythm and the Foreign Learner*, The Hague: Mouton.

Agnihotri, R.K. 1991. Sound patterns of Indian English: a sociolinguistic perspective. In: *English in India: Issues and Problems*, R.S. Gupta and K. Kapoor (eds.), 175–188, Delhi: Academic Foundations.

Agnihotri, R.K. 1999. India: English in India. In: *The Yearbook of South Asian Languages and Linguistics*, R. Singh (ed.), 184–198, New Delhi: Sage Publications.

Agnihotri, R.K. and Singh, R. (eds.). 2012. *Indian English: Towards a New Paradigm*, New Delhi: Orient Black Swan.

Allardyce, A. 1877. *The City of Sunshine. A Novel*, vol.1, William Blackwood.

Annamalai, E. 2008. Contexts of multilingualism. In: *Language in South Asia*, B.B. Kachru, Y. Kachru, and S.N. Sridhar (eds.), 223–234, Cambridge: Cambridge University Press.

Appa Rao, M. 1978. Orthography as underlying representation: a study of the vocalic phonology of R.P. and Telugu English within the framework of generative phonology. MLitt Dissertation, CIE&FL, Hyderabad.

Auer, P. 1993. Is a rhythm-based typology possible? A study of the role of prosody in phonological typology. KontRI Working Paper 21, Universität Hamburg.

Babu, P. 1971. Prosodic features in Indian English: stress, rhythm and intonation. Research diploma dissertation, CIE&FL, Hyderabad.

Balasubramanian, T. 1972. The vowels of Tamil and English: a study in contrast. *CIE&FL Bulletin* 9: 27–34.

Balasubramanian, T. 1975. Aspiration of voiceless stops in Tamil and English: an instrumental investigation. *CIE&FL Newsletter* 14–18.

Balasubramanian, T. 1980. Timing in Tamil. *Journal of Phonetics* 8: 449–468.

Balasubramanian, T. 2000. *A Textbook of English Phonetics for Indian Students*, Bombay: Macmillan Publishers India.

Bansal, R.K. 1976 [1969]. *The Intelligibility of Indian English*, 2nd edition, Monograph 4, Hyderabad: CIE&FL.

Bansal, R.K. 1983. *Studies in Phonetics and Spoken English*, Monograph 10, Hyderabad: CIE&FL.

Bansal, R.K. 1990. The pronunciation of English in India. In: *Studies in the Pronunciation of English: A Commemorative Volume in Honour of A.C. Gimson*, S. Ramsaran (ed.), 219–230, London: Routledge.

Bansal, R.K. and Harrison, J.B. 1974. *Spoken English for India*, New Delhi: Orient Longman.

Beckman, M. 1996. *Stress and Non-stress Accent*, Dordrecht: Foris.

Bhatia, T.K. and Baumgardner, R.J. 2008. Language in media and advertising. In: *Language in South Asia*, B.B. Kachru, Y. Kachru, and S.N. Sridhar (eds.), 377–394, Cambridge: Cambridge University Press.

Bingham, K. 1998. The intonation of Hindi function words. MA thesis, University of Texas, Austin.

Chaudhury, S. 1984. Some aspects of the phonology of Indian English. PhD dissertation, CIE&FL, Hyderabad.

CIEFL. 1972. *The Sound System of Indian English*, Monograph 7, Hyderabad: CIE&FL.

Coelho, G. 1997. Anglo-Indian English: a nativized variety of Indian English. *Language and Society* 26, 561–589.

Dauer, R. 1983. Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics* 11: 51–62.

Dauer, R. 1987. Phonetic and phonological components of language rhythm. Paper presented at *The XIth International Congress of Phonetic Sciences*, Tallinn.

Dhamija, P.V. 1976. A phonological description of Rajasthani English. M.Litt dissertation, CIE&FL, Hyderabad.

Dobson, E. 1968. *English Pronunciation 1500–1700*, Oxford: Oxford University Press.

Dyrud, L. 2001. Hindi-Urdu: stress-accent or non-stress accent? MA thesis, University of North Dakota, Grand Forks.

Féry, C., Pandey, P., and Kentner, G. 2013. The prosody of focus and givenness in Hindi and Indian English.

Gargesh, R. 2004. Indian English: phonology. In: *A Handbook of Varieties of English, vol. 1: Phonology*, W. Edgar, E. Schneider, and B. Kortmann (eds.), 993–1002, Berlin: Mouton de Gruyter.

Gargesh, R. 2006. Language issues in the context of Higher Education in India. Paper presented at the symposium on *Language Issues in English-Medium Universities across Asia*, University of Hong Kong, 8–9 June.

Gimson, A.C. 1962. Introduction. In: *English Pronouncing Dictionary*, London: Edward Arnold.

Goffin, R.C. 1934. *Some Notes on Indian English*, S.P.E. Tract No.41, 20–32, Oxford: Clarendon Press.

Gopalakrishnan, R. 2011 [2008]. *Second Language Phonology: Investigating Tamil English*, Saarbrücken: LAP Lambert Publishing Company. [MPhil Dissertation, Jawaharlal Nehru University, New Delhi.]

Gumperz, J.J. 1982. *Discourse Strategies*, Cambridge: Cambridge University Press.

Gupta, K.L. 1982. Dogri English: a phonological study. MLitt dissertation, CIE&FL, Hyderabad.

Harnsberger, J. 1994. Towards an intonational phonology of Hindi. Unpublished dissertation, University of Wisconsin, Madison.

Harnsberger, J.D. 1999. The role of metrical structure in Hindi intonation. Paper presented at the South Asian Language Analysis Roundtable, University of Illinois, Champaign-Urbana.

Hayes, B. 2004. The phonology of rhythm in English. *Linguistic Inquiry* 15: 33–74.

Hayes, B. and Lahiri, A. 1991. Bengali intonational phonology. *Natural Language and Linguistic Theory* 9: 47–96.

Hosali, P. 2000. *Butler English: Form and Function*, New Delhi: BR Publishing Corporation.

Hosali, P. and Aitchinson, J. 1986. Butler English: a minimal pidgin? *Journal of Pidgin and Creole Languages* 1 11: 51–79.

Indira, K. 2009. A spectrographic study using Praat of the phonemic system in the English of educated Yerukula speakers. MPhil dissertation, CIE&FL, Hyderabad.

Joseph, K.G. 1984. *Intonation Patterns of Telugu English*. MLitt dissertation, CIE&FL, Hyderabad.

Kachru, B.B. 1983. *The Indianization of English: the English language in India*, Oxford: Oxford University Press.

Kachru, B.B. 1996. *The Other Tongue: English across Cultures*, Delhi: Oxford University Press.

Keane, E. 2005. Prominence in Tamil. *Journal of the International Phonetic Association* 36: 1–20.

Keane, E. 2007. Investigating the intonational phonology of Tamil. ICPhS 2007 Satellite Workshop. http://www.linguistics.ucla.edu/people/jun/Workshop2007ICPhS/Papers/Keane_handout.pdf.

Khan, N. 1974. A phonological analysis of English spoken by all India Radio newsreaders. Post-Graduate Research Diploma dissertation, CIE&FL, Hyderabad.

Kumar, R., Kataria, A., and Sofat, S. 2003. Building non-native pronunciation lexicon for English using a rule based approach. In: *International Conference on NLP*, Mysore.

Kumar, R., Gangadharaiah, R., Rao, S., Prahallad, K., Rosé, C.P., and Black, A.W. 2007. Building a better IndEvoice using "More Data". In: *6th ISCA Workshop on Speech Synthesis*, Bonn, Germany.

Ladefoged, P. 1993. *A Course in Phonetics*, Fort Worth: Harcourt Brace College Publishers.

Latha, P. 1978. Intonation of Malayalam and Malayalee English: a study of comparison and contrast. MLitt thesis, CIE&FL, Hyderabad.

Maddieson, I. 1984. *Patterns of Sounds*, Cambridge: Cambridge University Press.

Madhavi, V. 2009. *A phonetic study of strong forms and weak forms in Telugu English.* MLitt dissertation, The English and Foreign Languages University, Hyderabad.

Mahesh, M. 2013. Rhythm and intonation in Malayalam. PhD thesis, Jawaharlal Nehru University, New Delhi.

Maxwell, O. and Fletcher, J. 2009. Acoustic and durational properties of IE vowels. *World Englishes* 28: 52–70.

Mesthrie, R and Bhatt, R. 2009. The study of new linguistic varieties. In: *World Englishes*, R. Mesthrie and R. Bhatt (eds.), Cambridge: Cambridge University Press.

Mohanan, K.P. and Mohanan, T. 1987. The lexical phonology of Malayali English: structure formation in transplanted second language systems. Unpublished manuscript, Stanford University.

Mullick, Y.J., Agrawal, S.S., Tayal,S., and Goswami, M. 2004. Text-to-phonemic transcription and parsing into monosyllables of English text. *Journal of the Acoustical Society of America* 115(5): 2544.

Nagarajan, H. 1985. Some Phonetic Features of Tamilian English. MLitt dissertation, CIE&FL, Hyderabad.

Nair, N.G. 1996. *Indian English Phonology: A Case Study of Malayalee English*, New Delhi: Prestige Books.

Nihalani, P., Tongue, R.K., Hosali, P., and Crowther, J. 2004. *Indian and British English. A Handbook of Usage and Pronunciation*, 2nd edition, Oxford: Oxford University Press.

Ohala, J.J. and Gilbert, J.B. 1979. Listeners' ability to identify languages by their prosody. In: *Probl|mes de Prosodie*,vol. II, P. Léon and M. Rossi, 123–131, Ottawa: Didier.

Pandey, P. 1980. *Stress in Hindustani English: A Generative Phonological Study*, Hyderabad: CIE&FL.

Pandey, P. 1981 [1994]. On a description of the phonology of Indian English. CIE&FL *Bulletin* 17: 11–19. Also in R.K. Agnihotri and A.L. Khanna (eds.), *Second Language Acquisition: Socio-cultural and Linguistic Aspects of English in India*, 198–207, New Delhi: Sage Publications.

Pandey, P. 1985. Word accentuation in Hindustani English in relation to Hindustani and English. PhD dissertation, University of Poona, Pune.

Pandey, P. 1989. Word accentuation in Hindi. *Lingua* 77: 37–73.

Pandey, P. 1990. Hindi schwa deletion. *Lingua* 82: 277–311.

Pandey, P. 2014. *Sounds and Their Patterns in Indic Languages*, New Delhi: Cambridge University Press India.

Pickering, L. 1999. An analysis of prosodic systems in the classroom discourse of native speaker and nonnative speaker TAs. PhD thesis, University of Florida, Gainesville.

Pickering, L and Wiltshire, C. 2000. Pitch accent in Indian-English teaching discourse. *World Englishes* 19(2): 173–183.

Prieto, P. 2002. Coarticulation and stability effects in tonal clash contexts in Catalan. In: *SP-2002*, 587–590.

Rajendran, S. and Yagnanarayana, B. 1996. Word boundary hypothesization for continunous speech in Hindi based on F0 patterns. *Speech Communication* 18, 21–46.

Rao, K.S.N. 1961. A footnote to the Indian pronunciation of the initial /k,t,p/ and /v/ and /w/ in English. *Indian Linguistics* 22: 160–163.

Ravisankar, G. 1994. *Intonation Patterns in Tamil*, Pondicherry: Pondicherry Institute of Linguistics and Culture.

Roach, P. 1982. On the distinction between *stress-timed and syllable-timed* languages. In: *Linguistic Controversies*, D. Crystal (ed.), 73–79, London: Edward Arnold.

Sahgal, A. and Agnihotri, R.K. 1988. Indian English phonology: a sociolinguistic perspective. *English World-Wide* 9: 51–64.

Schiering, R., Bickel, B., and Hildebrandt, K.A. 2012. Stress-timed = word-based? Testing a hypothesis in prosodic typology. *STUF – Language Typology and Universals* 65(2): 157–168.

Sen, A. 2003. Pronunciation rules for IE Text-to-Speech system. In: *ISCA Workshop on Spoken Language Processing*, Mumbai.

Sen, A. and Samudravijaya, K. 2002. Indian accent Text-to-Speech system for web browsing. *Sadhana* 27(Part I): 113–126 (February Issue).

Sethi , J. 1976. *English Spoken by Educated Punjabi Speakers in India: A Phonological Study*, Chandigarh: Punjabi University.

Sethi , J. 1978. The vowel system in educated Punjabi speakers' English. *Bulletin of the Central Institute of English* 14(2): 35–48.

Sethi, J. 1980. Word accent in educated Punjabi speakers' English. *Bulletin of the Central Institute of English* 16: 35–48.

Shekhar, G.N. 1993. A study of accent and intonation in all India Radio newsreaders speech. MPhil dissertation, CIE&FL, Hyderabad.

Skopeteas, S., Fiedler, I., Hellmuth, S., Schwarz, A., Stoel, R., Fanselow, G.,

Féry, C., and Krifka, M. 2006. Questionnaire on information structure. *ISIS*, 4.

Subbarao, G.1954. *Indian Words in English: A Study in Indo-British Cultural and Linguistic Relations*. Oxford: Clarendon Press.

Verma, D. 2005. Some aspects of the prosodic phonology of narrow and contrastive focus in Hindi and Hindi-English. MPhil dissertation, CIE&FL, Hyderabad.

Vijaykrishnan, K.G. 1978. Stress in Tamilian English: a study within the framework of generative phonology. MLitt dissertation, CIE&FL, Hyderabad.

Wells, J.C. 1982. *Accents of English*, Cambridge: Cambridge University Press.

Wells, J.C. 2000. *Longman Pronouncing Dictionary*, London: Longman.

Wenk, B. and Wioland, F. 1982. Is French really syllable-timed? *Journal of Phonetics* 10: 193–216.

Whitworth, G.C.1907. *Indian English: An Examination Made by Indians in Writing English*, Letchworth: Garden City Press.

Wiltshire, C.R. 2005. The "Indian English" of Tibeto-Burman language speakers. *English World-Wide* 26: 275–300.

Wiltshire, C.R. In press. Emergence of the unmarked in Indian Englishes with different substrates. In: *The Oxford Handbook of World Englishes*, M. Filppula, J. Klemola, and D. Sharma, Oxford: Oxford University Press.

Wiltshire, C.R. and Harnsberger, J.D. 2006. The Influence of Gujarati and Tamil L1s on Indian English: a preliminary study. *World Englishes* 25: 91–104.

Wiltshire, C.R. and Moon, R. 2003. Phonetic stress in IE versus American English. *World Englishes* 22(3): 291–303.

Yule, H. and Burnell, A.C. 1996 [1886]. *Hobson–Jobson: The Anglo-Indian Dictionary*, London: Wordsworth Reference.

Zaire, L. and Pandey, P. In press. Koren. In: *Sounds and Their Patterns in Indic Languages*. New Delhi: Cambridge University Press India.

18 Pronunciation and World Englishes

CECIL L. NELSON AND
SEONG-YOON KANG

Introduction

Pronunciation has frequently received a lot of attention in observation, analysis, and pedagogy, perhaps because it is an immediately noticeable and salient language feature (see, for example, Levis 2005: 369). As B. Kachru (1986: 140) put it, "It is … the pronunciation of a speaker which provides an index to the variety of his speech, or to a variety within a variety …. [O]ne does not have to be initiated in phonetics or linguistics to identify, for example, a speaker of American, British, or Indian varieties of English." In face-to-face speech interactions, we get a good deal of information about who we are dealing with from the opening exchanges, and we recognize that the flow of this information is in both directions, alternating as the speaker–hearer roles do (Nelson 2011: 33–34). At its most fundamental levels of utility, pronunciation serves to convey words and phrases that are recognizable as such and that make sense in the context of the situation.

The school of thought referred to here, *world Englishes*, may be seen to have had its early public exposures in volumes edited by B. Kachru (1982) and by L.E. Smith (1981). Its most fundamental tenet is that the many varieties of the English language "belong" to their users. For example, Smith (1983a: 7) wrote:

> When any language becomes international in character, it cannot be bound to any one culture. A Thai doesn't need to sound like an American in order to use English well with a Filipino at an ASEAN meeting. A Japanese doesn't need an appreciation of a British lifestyle in order to use English in his business dealings with a Malaysian.

A cogent expression of the world Englishes position on pronunciation may be found in Strevens, who wrote this definition of Standard English just as "world

The Handbook of English Pronunciation, First Edition. Edited by Marnie Reed and John M. Levis.

Englishes" was becoming an established phrase and point of view on varieties of the language around the world. He wrote (1983: 88) that Standard English is:

> … a particular dialect of English, being the only non-localized dialect … which may be spoken with an unrestricted choice of accent.

Streven's presentation invites attentive reading. The phrase "a particular dialect" might be thought at first to be at odds with the immediately following characterization of Standard English (as Strevens is using the term here) as a "non-localized dialect". However, this is the essence of the world Englishes approach to variety. There are many dialects – varieties – of English, but all of them share enough grammatical, lexical, and phonological features to be recognized under that cover term. Strevens went on to explicate this position in some detail. Importantly for the present purpose, he noted that it is natural to observe and to think of grammar and lexicon (subsumed under his term "dialect") and accent (pronunciation) as occurring in normal pairings: "Thus, in Dorset, Dorset dialect and Dorset accent are used." And just to be thoroughly clear about it, Strevens noted that mixings of the two aspects of language do not naturally occur: "Kentucky dialect is never spoken with a Dorset accent" (1983: 89). Thus, anyone's English of recognizably international utility can be spoken with any accent.

This is a concise presentation of the world Englishes position. As soon as anyone admits that there *are* British, US, and Australian *varieties* of English, *inter alia*, there would seem to be no rational, defensible way to draw a line *based on* such distinctions between "good" and "bad", "right" and "wrong" pronunciation. There is no cogent basis for deciding who gets to make the division or who can put such a declaration into effect in English teaching and learning, not to mention acquisition, around the world (see, for example, Levis 2005).

Historically, those who felt they could, or even needed to, set up such a yes/no dichotomy appealed to the tattered but still lively "native/non-native" partitioning. This was, one supposes, a step away from an even older point of view to the effect that only English English was "proper" and that everyone else's – while it might have been regarded as "native" had the term been current – was not.

To regard another's English as "foreign" immediately calls sociolinguistic attention to participants and setting, Firth's *context of situation* (1964: 66). It requires a certain mindset to be able to travel to another country, look around, and remark inwardly or to a fellow traveler, "Look at all the foreigners!"

In any case, there was and remains an (at least) implicit stance that categorizes US and Australian Englishes as "native" but Singaporean and Indian as "non-native". If the definitions are cast in terms of world-demographical provenance, they can be made to work; in terms of what Strevens called *primary language* (1992: 36) they work much less well, if at all. The "non-native" notion must *a priori* make more sense to a functionally monolingual person than to a multilingual one – and we cannot allow ourselves to pass too lightly over the well-worn observation that it is strikingly the case that "native" English users tend to fall in the *mono-* group.

Ferguson's often-quoted, less often followed, admonition that "the whole mystique of native speaker and mother tongue should probably be quietly dropped

from the linguists' set of professional myths about language" (1982: vii) has not been universally or consistently adopted. One can hope, at least, that in the minds of world Englishes adherents, the distinction conveys less stigma than in the past. It would have to be acknowledged, for instance, that people who were born and raised in Bangalore, who carry on their adult lives there, and who have been using English for a wide variety of personal and livelihood functions for most of that time, should be regarded as native speakers of *Indian* English.

One of the co-founders of the organization International Association for World Englishes and of the journal *World Englishes* is Larry E. Smith, who insightfully and clearly developed the partitioning of language functioning and analysis (with close reference to English, but not necessarily limited to discussions of just this language) into Intelligibility, Comprehensibility, and Interpretability (e.g., Nelson 2011: 32–34; Smith 1992: 76; Kachru and Smith 2008: 61–64). Intelligibility in this narrowed, technical sense is concerned with phonetics and phonology – pronunciation. Smith brought to the conscious awareness of the applied linguistics fields the necessity of recognizing that the question "Is this word or stretch of speech intelligible?" is not reasonable or answerable. A user of English is *found* to be intelligible *by* another user *in* a given context, perhaps taking into account mutual *familiarity with topic*, certainly involving each participant's degree of *familiarity with the pronunciations of others* in the immediate situation. Smith (1992) and Smith and Rafiqzad (1983) wrote these two telling summative statements:

> Our speech … in English needs to be intelligible only to those with whom we wish to communicate in English.
>
> (Smith 1992: 75)

> Since native speaker phonology doesn't appear to be more intelligible than non-native phonology, there seems to be no reason to insist that the performance target in the English classroom be a native speaker.
>
> (Smith and Rafiqzad 1983: 57)

Those two assertions capture the basic stance of world Englishes regarding pronunciation. The English-using world, which perhaps constitutes as much as a quarter of the Earth's population, according to Kachru and Smith (2008), is variety, not sameness, not conformity to external models (with some caution as to that last in EFL regions, where the distinction from ESL may still be found applicable and informative). Like the traditional EFL/ESL distinction, no part of the definitions of the Circles have ever appealed to alleged individuals' or groups' degrees of *proficiency* in the language.

One can look almost anywhere in the current world Englishes literature and find an expression of this view. For example, Sharbawi (2012: 179) writes:

> The current acoustic investigation of vowel contrasts in [Brunei English] is motivated by a few factors. First, the results of earlier studies have compared the findings to those of [British English]. It has since been realized that the practice of comparing the

> vowels of a particular English variety to those of an Inner Circle variety, for example, [Singaporean English] with [British English] … can be problematic because there is sometimes a tendency to view the phonological system of the new English variety as deficit [sic] and its features as erroneous. An alternative approach [i.e., the one adopted in this study] is to see a descriptive account of a vowel system independently without resorting to comparison with an Inner Circle English.

The world Englishes replacement for the constructs *native* and *non-native* has been, since 1985, the Three Circles model first presented by Kachru (1985). While some scholars have challenged this conceptualization on various grounds (Yano 2009), the historical bases of the global spread of English are difficult to deny. (See Kachru 2005: 211–220, a section entitled "On getting the Three Circles Model backwards", for a general rebuttal of such objections.)

The Inner-Circle English-using countries are those in which English, chronologically speaking, was first the primary language of majority populations of users for virtually all private and public functions: England, the US, Canada, Australia, and New Zealand. The Outer Circle comprises "transplanted English in new linguistic, cultural and social contexts", including countries in Africa, Asia, and South Asia, such as Nigeria, Singapore, and India. Expanding Circle varieties are those in, for example, Russia, Taiwan, and Korea, where English was neither transported as a primary language, as in the Inner Circle, nor as a colonial-era language, as in the Outer Circle. In the Expanding Circle, exhaustive listing of countries becomes problematic if not all but impossible. (See Kachru, Kachru, and Nelson 2006: 2–3; see also, for example, Kachru 1985 and Bautista and Gonzalez 2006: 130 and *passim*.)

These groupings are related to the concepts *norm-providing* and *norm-accepting* varieties (Kachru 1986: 84, 86–88). That is, American English speech communities provide their own norms of use, while to the extent that some other population of users, e.g., Koreans, rely on US English for their notions of correctness, they may be said to be norm-accepting. Based on their respective pronunciations, one of the present authors would be easily identifiable by any other English user as an American; the other, as Korean, or anyway as *not*-American. Even in a norm-accepting situation (likely more *de jure* than *de facto*), practicalities of transmission and learning/acquisition will keep pronunciations more or less distinct across populations of speakers.

This issue touches on the topic of identity, and it may be said that using a given variety of English can affect one's own self-image and the perceptions of others about a person or a group of users in two ways. First of all, the use of English in the Outer and Expanding Circles probably "sends signals" of modernity, higher education, and social and professional mobility (see, for example, Bautista and Gonzalez 2006: 131; Bolton 2006: 292; Kachru 1992a: 6). At the same time, the choice of English over another available language may be problematic in ways that would not readily occur to an Inner Circle speaker. That is, English may convey solidarity with or separateness from conversation participants, depending on *their* uses of English and their evaluations of such use. These parameters are not, of course,

always transparent, so making the linguistic choice may involve a degree of risk (see, for example, Kachru 1992b: 60–61 and 66–68, and Bamgbọṣe 1992).

A second consideration involves the choice of model or norm for one's (or a group's) English. Speakers may feel the push toward and the pull away from a particular variety of English, or find (perhaps have it pointed out to them) that they are accommodating their speech styles unconsciously (see, for example, Kachru 1992b: 57, and Shaw 1983). "Do I want to sound like a Korean or like an American?" is not a question that is likely to occur to an Inner Circle speaker, but it may be a vexed question for people in some parts of the English-using world.

An approach put forward by adherents of the English as a Lingua Franca (ELF) school of thought (e.g., Jenkins 2007) seeks another sort of solution to the cross-variety intelligibility issue. While Jenkins (2007: 2) writes that ELF is "an emerging English that exists in its *own right* and which is being described in its *own terms*" (emphases in original), there are no descriptions that would indicate that the pronunciations of all the varieties of English that would constitute ELF are trending toward similarity, let alone identity. In fact, it is the function of Jenkins' Lingua Franca Core (LFC) of pronunciation features to have ELF users acquire or learn a devised, recommended system of those elements (Jenkins 2007: 24 and elsewhere; Jenkins 2000: ch. 6 "Pedagogic priorities I: identifying the phonological core", 123–163; Jenkins 2009, 147–148). This hands-on, prescriptive adoption of a particular set of pronunciation recommendations for a subset of the world's English users is at odds with the descriptive view of self-normative English varieties that is presented in the world Englishes literature (see, for example, Nelson 2012; Kachru and Smith 2008: 2, 10 (f.n.), 84).

For clarification of these and other issues of pronunciation, we may turn to a frequently cited but perhaps less studied Asian variety of English, that of South Korea (officially, the Republic of Korea, hereafter Korea). Korea made a strenuous effort to keep ethnolinguistic homogeneity to build up national power and to keep its society stabilized in spite of serious contacts with other foreign languages such as Chinese and Japanese (Coulmas 1999), and its language played an important role in resistance against Chinese dominance and influence and Japanese imperialism under diplomatic, political, economic, or academic pressures (Coulmas 1999; Kaplan and Baldauf 2003). Surprisingly, however, Korea, an Expanding Circle English-using country, made an exception to English that has taken deep root in society and has affected the language used in all aspects of leisure, advertising, entertainment, business, education, mass media, and government over the past six decades (Chang 2008), where English is considered a language of opportunity for social and economic upward mobility and a representation of high social status and economic power, as it is in other Outer and Expanding Circle countries (Choi 2008; Ross 2008).

A revolutionary transition from grammar-translation instruction to communicative language teaching was brought about by Korea's hosting the 1986 Asian Games and the 1988 Summer Olympic Games (Baik 1992; Shim 1994); these two international athletic events acted as a catalyst for a necessity of fluency-oriented

English education. A national globalization project, *Segyehwa*, for national competitiveness in the worldwide economy, accelerated the need for a high level of oral communication skills and native-like, accent-free, pronunciation (Kaplan and Baldauf 2003; Shim and Baik 2000; Kim 2007), and accordingly, a plethora of language policies and proposals have been planned, discussed, and implemented, such as recruiting a large number of native English instructors (Chang 2009), building English-only zones and residential villages (Park 2009), learning English as a mandatory subject in elementary school (Park 2004; Shim and Baik 2000), shifting from grammar-translation instruction to a communicative English curriculum through the national curriculum reforms (Shin 2007), and discussing the possibility of enacting English as an official language in Korea (Yoo 2005).

Korean speakers of English are easily recognized as *not* Inner Circle speakers by their pronunciation, whether in their practised or unrehearsed speech, due to the influences of Korean phonological structure and process and speech styles, particularly in stress and intonation. Different from English, a stress-timed language, Korean is a syllable-timed language and so has a narrower pitch range, which makes Koreans' English sound – to users of other Englishes – flat, monotonous, lacking in rhythm, or exhibiting misplaced lexical and phrase stresses (Lee 2001). Above all, the Korean articulatory system makes Korean speakers of English conspicuous among speakers of other varieties. For example, English vowels have more fine-grained distinctions between front and back, between high and low, and between tense and lax than those of Korean, let alone English having more vowels than Korean (Cho and Park 2006). Therefore, Korean speakers of English have difficulty in apprehending and producing English vowels because of the phonological interference of the Korean sound system. For example, Koreans are apt to pronounce /æ/ (*hat*, *apple)* as /ɛ/ or /e/ since there is no /æ/ in Korean. Koreans do not distinguish short (lax) /ɪ/ (*sit*, *it*) from long (tense) /i/ (*seat*, *eat*) and typically do not distinguish between /ɜ:/ (*work*) and /ɔ:/ (*walk*) since there is no Korean equivalent for /ɜ:/ (Lee 2001). These are a few of the more salient vocalic features of a "Korean accent".

Even more distinctive characteristics of Koreans' English can be found in their pronunciation of consonants. For instance, some English consonants that do not exist in Korean, such as /f/, /v/, /ð/, /θ/ (*few*, *very*, *they*, *think*), are substituted and often pronounced as /p/, /b/, /d/, /s/ (*pew*, *berry*, *day*, *sink*), respectively. In particular, English /r/, a sound that does not exist in Korean, is considered by Korean speakers of English to be the most difficult, confusing, and noticeable to users of other English varieties (Lee 2001; Sung 2007); therefore, English /r/ (*rice*, *read*, *road*) is often pronounced (or perceived by hearers) as /l/ (*lice*, *lead*, *load*). Similarly, aspirated /p^h, t^h, k^h/ (*pill*, *till*, *kill*) may sound a little stiffer or stronger, similar to Korean tense unaspirated /p', t', k'/ under the influence of pronouncing Korean lax, tense unaspirated, and heavily aspirated consonants (Goddard 2005).

Almost all Korean learners and teachers of English aim at speaking English like Americans, consciously and unconsciously accepting norms of American English as the absolute canon, although these days there are other competing varieties of Inner Circle models, say, British or Australian English (Gibb 1999; Jung 2005; Shim 2002;

Yook 2006). Accordingly, American-*like* pronunciation has become a criterion of advanced proficiency because of the perceived need for a means of worldwide communication and country-external interactions, just as reading comprehension used to be the yardstick for successful foreign language learning in the past, when English was taught mainly for reading and writing through the grammar-translation method. Furthermore, due to the international dominance and socioeconomic power of the United States, pronunciation of *beoteo-balrin* ("buttered") American English is considered refined, is highly preferred over other varieties (Gibb 1999), and should be taught and learned in Korean English education (Yook 2005). Even an extreme misguided surgical operation, which is called 'linguistic surgery,' was developed to help Koreans pronounce native-like English sounds (J. Park 2009; Shin 2004). On the other hand, nativized English pronunciation (so-called Konglish) or Korean-accented English is widespread in people's daily speech and words are sometimes, on purpose, pronounced like Korean words. In particular, nativized, *koreanish* pronunciation of English is widespread in Korean television shows, including situation comedies, sketch comedies, and standup comedy routines (Park 2004), and is still valued, especially among the older generation, who learned English only as an academic subject or written language, with little exposure to spoken English, or Inner Circle English users. However, this is less so for younger people, who have had more opportunities to directly interact with American English and who want to sound like Americans. Especially because of the socioeconomic and diplomatic power of the United States, let alone English being a language of wider communication for education, economy, and diplomacy across the world, American-*like* pronunciation will likely become more and more preferred and highly evaluated in Korean society despite the exposure to other varieties of English.

Thus, Korea is an example of a controversial English-expansion context. Korean English speakers are identifiable by their accents, so in that sense may be regarded as users of "Korean English". Adherence or aspiration to in-country or to external norms will be one of the major criteria in determining whether English in Korea is to be recognized (by Koreans and by outsiders) as nativized and acculturated, moving at least toward becoming an additional language, or whether it will continue to be regarded as a norm-accepting, learned language.

The worlds' Englishes have evolved their distinctive pronunciations under sociolinguistic conditions which may affect any language's development. Speakers seek to make their speech appropriate and effective to as wide a variety of other users as they may find desirable, while maintaining their own ethnic, national, regional, and personal identities.

REFERENCES

Baik, M.J. 1992. Language shift and identity in Korea. *Journal of Asian Pacific Communication* 3: 15–31.

Bamgbọṣe, A. 1992. Standard Nigerian English: issues of identification. In: *The Other Tongue: English across Cultures*,

2nd edition, B.B. Kachru (ed.), 148–161, Urbana, IL: University of Illinois Press.
Bautista, M.L.S. and Gonzalez, A.B. 2006. Southeast Asian Englishes. In: *The Handbook of World Englishes*, B.B. Kachru, Y. Kachru, and C.L. Nelson (eds.), 130–144. Malden, MA: Blackwell.
Bolton, K. 2006. Varieties of World Englishes. In: *The Handbook of World Englishes*, B.B. Kachru, Y. Kachru, and C.L. Nelson (eds.), 289–312, Malden, MA: Blackwell.
Chang, K.-S. 2008. Status and function of English as a language of international/intercultural communication in Korea. *Organisation for Economic Co-operation and Development*. Retrieved from: http://www.oecd.org/edu/ceri/41506521.pdf.
Chang, B.-M. 2009. Korea's English Education Policy innovations to lead the nation into the globalized world. *Pan-Pacific Association of Applied Linguistics* 13: 83–97.
Cho, J. and Park, H.-K. 2006. A comparative analysis of Korean-English phonological structures and processes for pronunciation pedagogy in interpretation training. *Meta: Translators' Journal* 51: 229–246. doi: 10.7202/013253ar.
Choi, I.-C. 2008. The impact of EFL testing on EFL education in Korea. *Language Testing* 25: 39–62. doi: 10.1177/0265532207083744.
Coulmas, F. 1999. The Far East. In: *Handbook of Language and Ethnic Identity*, J.A. Fishman, 399–413, New York: Oxford University Press.
Ferguson, C.A. 1982. Foreword. In: *The Other Tongue: English across Cultures*, B.B. Kachru (ed.), vii–xi, Urbana, IL: University of Illinois Press.
Firth, J.R. 1964. On sociological linguistics. In: *Language in Culture and Society: A Reader in Linguistics and Anthropology*, D. Hymes (ed.), 66–70, New York: Harper & Row. (Reprinted from The technique of semantics. *Transactions of the Philological Society* 1935, 36–72.)
Gibb, M. 1999. A comparative study of attitudes towards varieties of English held by professionals and tertiary level students in Korea. *The Korea TESOL Journal* 2: 31–42.
Goddard, C. 2005. *The Languages of East and Southeast Asia: An Introduction*, Oxford: Oxford University Press.
Jenkins, J. 2000. *The Phonology of English as an International Language*, Oxford: Oxford University Press.
Jenkins, J. 2007. *English as a Lingua Franca: Attitude and Identity*, Oxford: Oxford University Press.
Jenkins, J. 2009. *World Englishes: A Resource Book for Students*, 2nd edition, New York: Routledge.
Jung, W.-H. 2005. Attitudes of Korean EFL learners towards varieties of English. *English Teaching* 60: 239–260.
Kachru, B.B. (ed.). 1982. *The Other Tongue: English across Cultures*, Urbana, IL: University of Illinois Press.
Kachru, B.B. 1985. Standards, codification and sociolinguistic realism: the English language in the Outer Circle. In: *English in the World: Teaching and Learning the Language and Literatures*, R. Quirk and H.G. Widdowson (eds.), 11–30, Cambridge: Cambridge University Press.
Kachru, B.B. 1986. *The Alchemy of English: The Spread, Functions and Models of Non-native Englishes*, New York: Pergamon Press.
Kachru, B.B. 1992a. Introduction: the other side of English and the 1990s. In: *The Other Tongue: English across Cultures*, 2nd edition, B.B. Kachru (ed.), 1–15, Urbana, IL: University of Illinois Press.
Kachru, B.B. 1992b. Models for non-native Englishes. In: *The Other Tongue: English across Cultures*, 2nd edition, B.B. Kachru, 48–74. Urbana, IL: University of Illinois Press.
Kachru, B.B. 2005. *Asian Englishes: Beyond the Canon*, Hong Kong: Hong Kong University Press.
Kachru, B.B., Kachru, Y., and Nelson, C.L. 2006. Introduction: the world of World

Englishes. In: *The Handbook of World Englishes*, B.B. Kachru, Y. Kachru, and C.L. Nelson (eds.), 1–16, Malden, MA: Blackwell.

Kachru, Y. and Smith, L.E. 2008. *Cultures, Contexts, and World Englishes*, New York: Routledge.

Kaplan, R.B. and Baldauf, R.B. 2003. *Language Planning from Practice to Theory*, Clevedon, UK: Multilingual Matters.

Kim, Y.S. 2007. Korean adults' attitudes towards varieties of English. Unpublished Master's thesis, University of Edinburgh. Retrieved from: https://www.era.lib.ed.ac.uk/bitstream/1842/1903/6/Kim%20YS%20diss%2007.pdf.

Lee, J.-A. 2001. Korean speakers. In: *Learner English: A Teacher's Guide to Interference and Other Problems*, 2nd edition, M. Swan and B. Smith (eds.), 325–342, Cambridge: Cambridge University Press.

Levis, J. 2005. Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly* 39: 369–377. doi: 10.2307/3588485.

Nelson, C.L. 2011. *Intelligibility in World Englishes*, New York: Routledge.

Nelson, C.L. 2012. Review of *English as a Lingua Franca: Attitude and Identity* by Jennifer Jenkins. *World Englishes* 31(1): 139–141.

Park, J.S.-Y. 2004. "Baby, Darling, Honey!" Constructing a competence of English in South Korean TV Shows. *Texas Linguistic Forum* 47: 143–154.

Park, J.-K. 2009. English fever in South Korea: its history and symptoms. *English Today* 97(25), 50–57. doi: 10.1017/S026607840900008X.

Ross, S.J. 2008. Language testing in Asia: evolution, innovation, and policy challenges. *Language Testing* 25: 5–13. doi: 10.1177/0265532207083741.

Sharbawi, S. 2012. Revisiting the vowels of Brunei English. *World Englishes* 31: 177–197. doi: 10.1111/j.1467-971X.2011.01725.x.

Shaw, W.D. 1983. Asian student attitudes towards English. In: *Readings in English as an International Language*, L.E. Smith (ed.), 21–34, Oxford: Pergamon. (Originally published in *English for Cross-Cultural Communication*, L.E. Smith (ed.), 1981.)

Shim, R.J. 1994. Englishized Korean: structure, status, and attitudes. *World Englishes* 13: 225–244. doi: 10.1111/j.1467-971X.1994.tb00309.x.

Shim, R.J. 2002. Changing attitudes towards TEWOL in Korea. *Journal of Asian Pacific Communication* 12(1): 143–158. doi: 10.1075/japc.12.1.09shi.

Shim, R.J. and Baik, M.J. 2000. Korea (South and North). In: *Language Policies and Language Education: The Impact in East Asian Countries in the Next Decade*, W.K. Ho and R.Y.L. Wong (eds.), 173–196, Singapore: Times Media.

Shin, H. 2007. English language teaching in Korea: toward globalization or localization. In: *The Handbook of English Language Teaching*, J. Cummins and C. Davison (eds.), 75–86, Oxford: Wiley-Blackwell.

Smith, L.E. (ed.). 1981. *English for Cross-Cultural Communication*, New York: St. Martin's Press.

Smith, L.E. 1983a. English as an international language: no room for linguistic chauvinism. In: *Readings in English as an International Language*, L.E. Smith (ed.), 7–11, Oxford: Pergamon.

Smith, L.E. (ed.). 1983b. *Readings in English as an International Language*, Oxford: Pergamon.

Smith, L.E. 1992. Spread of English and issues of intelligibility. In: *The Other Tongue: English across Cultures*, 2nd edition, B.B. Kachru (ed.), 75–90, Urbana, IL: University of Illinois Press.

Smith, L.E. and Rafiqzad, K. 1983. English for cross-Cultural communication: the question of intelligibility. In: *Readings in English as an International Language*, L.E. Smith (ed.), 49–58, Oxford: Pergamon. (Originally published in *TESOL Quarterly* 13: 371–380. doi: 10.2307/3585884.)

Strevens, P. 1983. What is "Standard English"? In: *Readings in English as an International Language*, L.E. Smith (ed.), 87–93. (Originally published in *RELC Journal* 1982. doi: 10.1177/003368828101200201.)

Strevens, P. 1992. English as an international language: directions in the 1990s. In: *The Other Tongue: English across Cultures*, 2nd edition, B.B. Kachru (ed.), 27–47, Urbana, IL: University of Illinois Press.

Sung, E.-K. 2007. The relationship between phonetics and phonology: problematic cases for a unified model. *Studies in Phonetics, Phonology and Morphology* 13: 295–313.

Yano, Y. 2009. The future of English: beyond the Kachruvian Three Circle Model? In: *Global Englishes in Asian Contexts: Current and Future Debates*, K. Murata and J. Jenkins (eds.), 208–225, New York: Palgrave Macmillan.

Yoo, O.K. 2005. Discourse of English as an official language in a monolingual society: the case of South Korea. *Second Language Studies* 23: 1–44.

Yook, C. 2006. Korean University students' attitudes toward varieties of English. In: *Proceedings of the Joint Conference of AAAL and ACLA/CAAL 2006*, Montreal, Canada.

Part V Pronunciation and Language Acquisition

19 Acquisition of the English Sound System

MARILYN MAY VIHMAN

Early studies of child language

The earliest publications to address phonological development were diary studies by European scholars. These culminated in Jakobson's attempt to build a grand model of the "universal and constant laws" that might govern the process (Jakobson 1949: 378). English played only a small part in these theoretical beginnings. However, in the past 40 years of intensive acquisition research inspired by Chomsky's (1965) strong nativist claims, data from children acquiring English have heavily dominated the field. This makes it particularly interesting to ask what the specific characteristics of English phonology are from a developmental point of view, since English has implicitly served as a kind of general model for acquisition (see the "universal tendencies … or constraints" proposed by Smith (1973: 206) on the basis of his generative-rule-based study of his son Amahl's acquisition of English).

Fortunately, cross-linguistic studies of both perceptual processing and early word production have become so much more common in the past 10 or 20 years that it is now possible to place the acquisition of English in a broader framework, in which the pervasive individual differences across children can be weighed against the typological evidence to identify those aspects of the ambient language that most clearly affect early infant language development. At the same time, such a framework allows us to separate out the "universal" elements (like those that concerned Jakobson, still embedded in markedness ideas today: see Kager 1999; Kager, Pater, and Zonneveld 2004, but with the advantage of a far more extensive database than was available earlier). It also allows us to consider the patterns of English in relation to perceptual and motoric aspects of infant development more generally.

The Handbook of English Pronunciation, First Edition. Edited by Marnie Reed and John M. Levis.

Rhythm class: the first ambient-language influence

English was long taken as the model "stress-timed" language, classically contrasted with the "syllable-timing" of languages like French or Spanish (Pike 1945; Abercrombie 1967). However, empirical studies have failed to identify any solid basis for this persistent two-way typology (see Dauer 1983). More recently, an approach to quantifying rhythm class along a continuum has been widely adopted instead (Ramus, Nespor, and Mehler 1999; Grabe and Low 2002; White and Mattys 2007 – but see now Arvaniti's (2012) thorough-going questioning of these methods). English continues to serve as the most characteristic language at the stress-timing end of the continuum.

This characterization of English is relevant here because infants show, from birth, a sensitivity to native-language rhythms, grounded in their pre-natal auditory experience with the sound of speech as filtered through the amniotic surround in the last trimester, when the auditory system is complete (Lecanuet 1993), preferring that language in experimental studies (Cooper and Aslin 1994; Mehler et al. 1988) and also distinguishing non-native languages, but only if they differ in rhythm class (Nazzi, Bertoncini, and Mehler 1998). Importantly, it is not prosodic differences alone that appear to support these infant responses but some combination of rhythm with other prosodic properties or with the characteristic phonotactic patterning of the language (Ramus 2002). The conclusion that we may draw from these studies is that the enduring characterization of English as "stress-timed" is best thought of as resulting not only from its lexically meaningful use of strong stress and concomitant vowel reduction but also from its inclusion of complex and varied syllables, with their codas, clusters, and diphthongs (Laver 1994: ch. 16.6). As we shall see, most of these elements are challenging to infant learners.

Infant speech perception: gaining knowledge of the native language

Along with the language-specific experience of rhythm, infants are well equipped, in the first months of life, to discriminate segmental contrasts. This can be seen as one of the biological foundations for language learning, although it is specific neither to language nor to humans (Jusczyk 1997; Vihman 2014). It is now well established, however, that this early ability fades quite rapidly with exposure to a particular language, resulting in infants already being more responsive by the end of the first year to the differences between phonemes contrasted in their own language than to unfamiliar contrasts (Werker and Tees 1984). The mechanism behind the phenomenon now known as "perceptual narrowing" (Lewkowicz 2011; Maurer and Werker 2014) remained unexplained for some 20 years. Since the mid-1990s, however, the importance of distributional or statistical learning in infancy has been intensively studied, mainly through the experimental use of artificial languages; this has led to the hypothesis that it is experience with the bimodal distribution of variants in the input (which results from the existence of a phonological contrast) that maintains in the infant listener the ability to discriminate, while

essentially unimodal (or unstructured) distribution of phones not supported by phonological contrast does not (Maye, Werker, and Gerken 2002; see also Anderson, Morgan, and White 2003).

Dramatic perception-related advances have been shown experimentally to occur in the first year, especially from 6 to 9 months. At the earlier age, infants exposed to English show a familiarity preference for listening to their own language only when the contrasting language is prosodically distinct. Specifically, American infants attend longer to an English word list when contrasted with Norwegian but not when contrasted with Dutch; by 9 months American infants listen longer to English in comparison with Dutch as well, demonstrating an advance in familiarity with the segmental level of speech at which the differences between English and Dutch become apparent (Jusczyk et al. 1993). Both the earlier preferential response based on the prosody of English and the later response based on the common segmental patterns of English are presumably the outcome of distributional learning based on consistent exposure to input that demonstrates these ambient language characteristics.

Further advances suggest the same kind of implicit learning. At 9 but not at 6 months, infants learning English prefer to listen to the more common strong–weak or trochaic pattern of English disyllabic words than to the less common weak–strong or iambic pattern (Jusczyk, Cutler, and Redanz 1993) and to common than to uncommon (but nevertheless permissible) phonotactic sequences (Jusczyk, Luce, and Charles-Luce 1994). Similarly, by 9 months infants distinguish commonly occurring within-word consonant clusters from those that occur only between words, at the same time demonstrating an expectation of word-initial stress that leads them to associate within-word clusters with the typical English strong–weak lexical pattern (Mattys et al. 1999). Further evidence of the effect of the dominant trochaic pattern is seen in word-learning experiments in which infants familiarized with trochaic nonwords recognize these in (or "segment" them out from) a short passage by age 7.5 months, whereas familiarization with iambic nonwords leads to segmentation only in infants three months older (Jusczyk, Houston, and Newsome 1999).

It has so far proven impossible to replicate this last study with infants as young as 7.5 months exposed to other languages. The ability to segment (unfamiliar) disyllabic words trained in the laboratory has been demonstrated for Dutch only by 9 months (Kuijpers et al. 1998) and for French only considerably later – by 12–16 months (Nazzi et al. 2006), although *monosyllables* familiarized in the laboratory are segmented by 8 months in English (Jusczyk and Aslin 1995), French (Gout 2001, as cited in Nazzi et al. 2006) and German (Höhle and Weissenborn 2003). Most strikingly, even infants exposed to British rather than American English have proven unable to recognize trained disyllabic words in passages in two British labs (DePaolis et al. 2012). Since in this case differences between the languages, or rather dialects, would seem insufficient to account for the failure to replicate the findings of Jusczyk, Houston, and Newsome (1999), the explanation seems likely to involve differences in the extent of prosodic modulation or "exaggeration" in speech to infants in the two cultures (see Fernald et al. 1989), an account that receives

further support from the fact that infants exposed to Canadian French – where North American cultural preferences for highly modulated "baby talk" may also be seen – show the familiarization effect for disyllables as early as American infants (Polka and Sundara 2012).

The experimental studies reviewed above provide clear evidence of advances, over the first year, in familiarity with the prosodic and segmental patterns of the ambient language. However, infants also begin to gain familiarity with the form of particular lexical items over this period. The very first word forms to be recognized, not surprisingly, are those that refer to the central characters in the infant's life – the infant himself (Mandel, Jusczyk, and Pisoni 1995) and his caretakers (Bortfeld et al. 2005). The evidence from American studies places knowledge of such names as early as four to six months although British studies have been unable to replicate the findings (Vihman and Keren-Portnoy 2013a).

In a separate line of study, infants have been found to show a robust ability to recognize *untrained* words familiar from everyday exposure and presented in word lists by 11 months, but not earlier (Vihman et al. 2004 (with British infants)). Manipulations of the forms of these common words have established that infant recognition is based particularly on the shape of the accented syllable – in English, on the word-initial consonant specifically (Vihman et al. 2004). In a follow-up study DePaolis, Vihman, and Keren-Portnoy (2014) found that the same common words could be recognized when embedded in sentences – i.e., could be *segmented*, *without* familiarization in the lab – only one month later, at 12 months. These studies test a different aspect of infant learning. Rather than demonstrating advances in implicit familiarity with the sound of the native language, they establish long-term infant memory for words that recur from one day to the next, in the routinized situations of the child's life. Thus it is not surprising that infants succeed in these studies a bit later – but it is also worth noting that the findings have been successfully replicated, with infants of the same age, wherever they have been tested (i.e., for isolated words: in Dutch, Swingley 2005; Italian, Vihman and Majorano 2014; and American English, DePaolis, Keren-Portnoy, and Vihman 2010; see also Vihman et al. 2007 for a replication using both Event-Related Potentials and the behavioral Head-turn Preference Procedure, testing cross-sectional groups of British infants at 9, 10, 11, and 12 months).

One aspect of development over the first year that we have not yet considered is production. It is striking that many of the changes we report above occur between 6 and 9 months – an age range that closely resembles that usually cited for the emergence of the first adult-like syllables, or "canonical babbling", in most typically developing infants (6–8 months in Oller 2000). These facts are likely to be related. Production of speech-like syllables provides the infant with cross-modal familiarity (internal, or proprioceptive, as well as external, or auditory, and, in the case of labials at least, also visual) with sound patterns that necessarily also occur in input speech, although the match will in most cases be only approximate (and will differ in characteristic ways between the adult male and female voices and that of the infant him- or herself; for a model of the way in which this difference may be overcome to allow recognition of the match see Callan et al. 2000). The

cross-modal experience should be a particularly potent aid to the infant in beginning to recognize words in the longer sequences to which he or she is primarily exposed (Bahrick, Lickliter, and Flom 2004) – i.e., in the segmentation task with which so many studies have been concerned.

Two recent studies were designed to test the proposal that infants' own vocal production influences the way they process speech. In both cases, infants were recorded in multiple home sessions until they showed frequent and stable use of one or more consonants (British English, 18 infants: DePaolis, Vihman, and Keren-Portnoy 2011; Italian, 30 infants: Majorano, Vihman, and DePaolis 2014; for a differently designed study of infants learning British English or Welsh, with similar results, see DePaolis, Vihman, and Nakai 2013). They were then tested in the lab with nonwords that featured a stop consonant that the infant was consistently producing (disregarding differences in voicing, which are not well controlled at this age), one that the infant was not yet producing with any regularity and a fricative pair (/s, z/) that none of the infants had in repertoire. The findings were the same in the two studies: infants who had achieved consistent production of only a single consonant preferred to listen to nonwords featuring that consonant, whereas infants with good production experience of at least two different consonants showed a significant preference for the *unknown* stop pair (the groups showed similar interest in the fricative pair, which was unrelated to production experience). The findings, though seemingly paradoxical, can be interpreted in terms of the hypothesis of a matching process between infant vocal production and input speech (an "articulatory filter" in Vihman 1993). Once a single adult-like consonant is part of an infant's regular production repertoire, that consonant, or more likely the syllables in which it occurs, gains particular salience. However, at the point when two or more such consonants are in repertoire, the infant begins to generalize, gaining a stronger sense of phonological possibilities and a concomitant interest in (or responsiveness to) the unfamiliar sounds (see Hunter and Ames 1988 for a general model of shifts in infant attention from what is familiar to what is novel, and Vihman, DePaolis, and Keren-Portnoy 2014, for further discussion of these findings).

First word production

Efforts to find ambient language effects through adult listeners making judgments as to infants' origins based on their babble have proven largely ineffective (Engstrand, Williams, and Lacerda 2003). However, close analysis of infant vocalizations provides good evidence of such effects already in the prelinguistic period. As could be expected, based on the findings from experimental studies of perceptual processing, *prosodic* aspects of the language of exposure are the first to be expressed in infant production. Whalen, Levitt, and Wang (1991) identified more rising pitch contours in French infants' reduplicated babbling than in those of English infants (age range 6 to 12 months). This agrees with Kent and Murray (1982), who also reported primarily falling contours for their American subjects over the first year. A study of the vowels produced by five 10-month-olds each

exposed to British English, French, Arabic, and Cantonese showed subtle differences within the low and central vowel space typical of this age, reflecting the patterning of vowels in the adult languages (Boysson-Bardies et al. 1989): English infants tended to produce more front vowels, in agreement with Kent and Murray (1982), reflecting the relatively high incidence of front vowels in adult English. With regards to consonants, a core set is consistently identified in the babbling of infants exposed to any language (Locke 1983), primarily stops and nasals, glottals, and glides. However, consonants also show an effect of ambient language influence as early as 10 months. Based on four groups of five infants each learning English, French, Japanese, and Swedish, for example, Boysson-Bardies and Vihman (1991) reported significantly more use of labials in English and French than in the other two languages.

What was suggested already in the first decade of audio-recorded observation of infant production (Oller et al. 1976) was later confirmed in studies of 10–20 infants acquiring American English: babbling practice is directly related to the first word forms of any given infant (Vihman et al. 1985; Vihman, Ferguson and Elbert 1986; McCune and Vihman 2001). Thus the tendencies that we see in babbling – in which a limited production repertoire constrains the range of possible ambient-language effects – are also seen in the first words, which tend to be not only similar to babbling but, at the same time, relatively accurate replicas of their adult models. The accuracy of the first word forms, first noted by Ferguson and Farwell (1975), can be explained in terms of the articulatory-filter proposal mentioned above. Practice, through babbling, with certain vocal patterns leads to deeper knowledge of those patterns, which accordingly become particularly salient to the infant in input speech. Given repeated exposure to certain high-frequency lexical items, the first words that an infant attempts are likely to be unconsciously "selected" from among those that match the sounds he or she is already able to make. The result is not only continuity with babble but also highly constrained first-word targets and relative accuracy in first word production. To illustrate these latter points the first 5–6 words of the 17 monolingual English-learners included in Appendix I in Menn and Vihman (2011), are reproduced here (see the Appendix in this chapter).

We can draw on this sample – a mix of diary and observational studies, with child ages ranging from 9 to 20 months (mean 12 months) – to gain a more concrete idea of the starting point for the acquisition of the English phonological system. About half (41) of the 83 first words attempted are monosyllables, with 40 disyllables and one instance each of *banana* and *patty-cake*, both produced as disyllables; this compares with a cross-linguistic mean of 32% monosyllables over all words attempted by the 48 children (Menn and Vihman 2011; the American children actually *produce* slightly more of the words as monosyllables: 0.55). For comparison, a mean of 0.69 of the content words produced by five American mothers in speech to their 12-month-olds were monosyllables (Vihman et al. 1994a: the mothers produced 0.23 disyllables and 0.08 longer words). Thus exposure to English input leads the American children to attempt and produce more monosyllabic words than is "universally" typical of first-word production.

The onsets in the Appendix are single consonants in all but eight of the targets (disregarding glottal stop), with just five words – *uh-oh* (3 occurrences), *up* (2), *all-gone*, *all done*, and *Edgar* – accounting for the remainder. Of the 75 onset consonants, all but 14 are stops, nasals, /h/, or glides (0.81 altogether), in accordance with the phonetic tendencies of babbling. A few words account for the exceptions (*that/there* (3), *juice*, *light*, and *see* (2 each) and five others). For comparison, the mothers sampled in Vihman et al. 1994a produced 0.56 initial stops and 0.11 nasals (0.23 fricatives or affricates, 0.17 liquids). The children's own forms match the target onset consonant in all but 20 words (disregarding both voicing changes and onset-vowel insertions), which include all of those with fricative, affricate, or liquid onsets. Finally, the single most commonly targeted onset consonant is /b/ (21 words), but coronal onsets are slightly more commonly targeted than labials (33 coronals, or 0.49, excluding the seven h-initial words; 31 labials, or 0.46); the velars are underrepresented, at 0.06 (compare the mothers' sample, with 0.34 initial labials, 0.43 coronals, and 0.22 velars). On the other hand, the labials *match the targets* in the child forms (except *banana*, reduced to its final syllables), while the coronals have varied outcomes.

Only four target words have onset clusters (*block*, *cracker*, *quack-quack*, *squirrel*); the children generally reduce the cluster to a stop, although the child attempting *cracker* variously produces it with [p-], [kw-], [w-], and [k-]. Including the mid-vowel off-glides as well as /ɑɪ/ and /ɑʊ/ (/ɔɪ/ is never targeted), 30 target words have diphthongs (12 different words). Of these, only four are never produced with a diphthong (*bow-wow*, *Jacob*, *nose*, *uh-oh*). Thus clusters are clearly more challenging than diphthongs.

Finally, consider two more aspects of these first words. Only 25 words with codas are targeted (0.30), compared with a mean of 0.67 of input content words with codas in English (Vihman et al. 1994b); of these, only four are produced with a coda consonant, all of them sibilants (*box*, *bus*, *juice*, *shoes*). Besides these four words, which all include stop onset as well as fricative coda, only six words have more than one true consonant (i.e., excluding combinations with glide or glottal) within a single word (*cracker*, *dog*, *doggie*, *Jacob*, *put on*, and *thank you*).

This then is the point of departure for acquisition of the English phonological system. The first words are close to their targets in length and in onset consonant. Child "selection" or bias in attempting words is apparent in the predominance of one- and two-syllable targets, although English content words provide relatively few challenging long words in any case (as compared with Italian, Japanese, or Spanish, for example). Similarly, the predominance of stop and nasal onsets in the words targeted seems to reflect infant preferences. There is also a bias in favour of /b/ and a clear advantage for labials in production. Clusters tend not to be targeted but diphthongs pose no apparent obstacle, although they are produced where required less than two-thirds of the time. Words with codas are undertargeted and the coda is seldom produced when needed. The additional, perhaps less obvious, difficulty for first-word production is presented by the need to remember, plan, and articulate two (or more) different consonants in a single-word production; this is seldom achieved at this point.

English word templates

We have suggested that most learning, in the prelinguistic period, is implicit, distributional (based on gaining familiarity with the prosodic and segmental patterns most frequently heard), and procedural (developing motoric routines that underlie repeated production of particular sounds and sequences of sounds). To learn to produce word forms in appropriate situations of use, however, the infant must draw on explicit learning as well – learning with attention and, eventually, with *intention,* often in dyadic interaction. This is the foundation for phonology, since the construction of a phonological system depends on word learning (Vihman and Keren-Portnoy 2013b).

Once a child has begun to produce a few words, he or she is in a position to learn from his or her own output – a small but highly familiar "database". As more new words are learned the infant's knowledge of the sound system continues to grow (and the receptive vocabulary is generally much larger than the expressive vocabulary); however, the child's repertoire of production plans – particularly for different consonants – grows far more slowly. Accordingly, we frequently see the child settle on a small number of prosodic structures or patterns that have been called "word templates"; this provides a "holding pattern", while the child's motoric and planning skills – and their memory for word forms – improves. The templates differ by child but show a "family resemblance" within language groups, so that we can look for characteristic patterns used by children acquiring English (Vihman, in press).

Early word templates for three of the children included in the Appendix have been described longitudinally, over the period in which their favored templates first developed. We draw here on those descriptions, which show considerable inter-child variability, and then compare these children with others, including two children whose prosodic structures have been described at a more advanced lexical point.

Vihman and Velleman (1989) detail the emergence of a template that seems to have been designed to allow Molly to produce codas, which she targeted frequently but nevertheless found difficult to produce. The study covers five months, from her first spontaneous use of four words in a 30-minute session (the start of established word use: the "four-word-point", or 4wp, at 10 months) to a cumulative vocabulary of over 70 words (35 words in the session). Both stop and nasal codas were attempted, through a sequence of identifiable stages – presystematic production, experimentation, and emergence of a predominant pattern or template. For both coda types, this first involved the addition of a support vowel (e.g., *bang* [pan:ə], *clock* [kak:ɪ̯], both at 1;1) and later the restructuring of target forms to fit the template (e.g., *Nicky* [ɪn:i], *glasses* [kak:^{h}i̯], both at 1;3).

Vihman et al. (1994b) recount the emergence over 6 to 8 months of templates in two children. Timmy produces CV and CVCV forms almost exclusively, sometimes with the addition of a nontarget onset vowel. His range of consonant use in word forms grows very gradually over the period from 10 months (4wp: [ba] only, with variants including both voiced and voiceless bilabial fricatives) to 16 months,

when Timmy has nine consonants and the three corner vowels, variously used with [b, t, k, m, n]. By this time Timmy is producing a limited number of variegated disyllables (e.g., *cookie* [kaki], *goodbye* [gaba]), some of them reflecting restructuring, similar to what we saw in Molly (*Simon* [nama], [nimi], *coffee* [kuki], *good boy* [kibi]).

The second child, Alice, showed unusually high use of the glide [j] in her babbling and first words. The word pattern she developed began with a high proportion of use of front-rising diphthongs (10 months), which was then paralleled by a preference for the disyllabic sequence <(C)VCi>, with palatalization often affecting the medial consonant (e.g., *blanket* [baji], *dolly*, *daddy* [daji], and, by 16 months, *belly* [vei], *bunny* [buɲ:i, beiŋji], and *shiny* [ta:ji] along with such more radically restructured forms as *elephant* [ʔæɪŋɨ, ʔai:njʌ], *flowers* [pa:ji], *iron* [ʔaɪŋ:, ãĩji], *lady* [jeiji, ijei], and *mommy* [ma:ɲi, əma:ɲi] (notice the focus here on targets ending in *-i,* a common English pattern for disyllables, especially in speech to children).

The most commonly reported pattern is probably consonant harmony (less used in English than in a more rhythmically regular language such as Finnish, however: Vihman and Wauquier, in press). An example is seen in Menn's (1971) diary study of her son Danny, whose first words appeared at 16 months and who developed, by 25 months, a strong harmonized <CVC> template (e.g., *bread* [bʌb], *jeep* [bip], *dog* [gɔg]). In contrast, Jaeger (1997) reports her daughter's more unusual use, by the time she had some 100 words (age 23 months), of a front–back consonant melody used in both one- and two-syllable words, sometimes with metathesis to achieve the favored structure (e.g., *butter* [pʌtu], *cheek* [tik^{h}], *frog* [pak^{h}], but also *David* [pita], *kite* [taɪk], and *sheep* [piç]; see also Vihman and Croft 2007).

Two classic studies of templates in children acquiring English illustrate additional patterns. Waterson (1971) describes several different "schemas" or prosodic structures into which her son organized his word forms at a time when he had some 150 words in use (aged 17 to 19 months). These include monosyllables with sibilant coda (*brush* [byʃ], *dish* [dɪʃ], *fetch*, *fish* [ɪʃ], *vest* [ʊʃ]) and disyllables with reduplication or harmony (*another* [ɲaɲa], *finger* [ɲɪ:ɲɪ]; *biscuit* [be:be:], *Bobby* [bæbu:]). Priestly (1977) describes his son's four-month use of a <CVjVC> pattern (another "melody"), at age 22 months, when he had well over 100 words. Here again some forms were relatively similar to the target (*peanut* [pijat], *carrot* [kajat]) while others freely restructured target forms (*chocolate* [kajak], *flannel* [fajan], *rhinocerous* [rajan]).

Harmony and melody patterns alike provide the child with support for planning as well as memory, in that a set frame with variable elements is more accessible for both purposes than a set of open choices. The very idiosyncracy of child templates makes it difficult to generalize from them, but these child "solutions" to the problem of remembering and producing a growing set of forms give us a good idea of what constitutes a challenge. As we see from these few examples, some templates address the problem of codas, others that of changing vowels or place or manner of consonants across the word; some deal with more than one of these issues.

English phonology at age two

To obtain "norms" indicating the consonant use to be expected at different ages, early studies used single-word naming tests based on picture presentation (e.g., Sander 1972). As Stoel-Gammon (1987) pointed out, this is not generally successful in arriving at an idea of two-year-old phonology, since many children of that age are resistant to testing and those able and willing to participate may not be representative of the age group as a whole. Accordingly, Stoel-Gammon used recordings of spontaneous speech to obtain data from a relatively large group of American children. Out of 34 participants all but one produced more than 10 different adult-based words in the session and were accordingly included in the study. The transcripts used for analysis were based on a maximum of 50 words; these variably reflected from 20 to 112 different word types (mean 36). This thus corresponds roughly to the period of word template use described above, although not all children necessarily make use of them (Vihman 2014: Appendix 3). Stoel-Gammon reports three analyses of her data: (i) word shapes produced, (ii) inventories of initial and final consonants, and (iii) accuracy (using Shriberg and Kwiatkowski's (1982) percent consonants correct [PCC] measure).

i. The monosyllabic word shapes CV and CVC occurred in virtually all samples. Disyllabic CVCV (which dominates the production of children learning many European languages) occurred in 26 samples (79%) and disyllabic words with codas CVCVC in 22 (67%). At least two different clusters occurred initially in 58% of the samples, finally in 48%, and medially in only 30%.
ii. Consonant inventories included a mean of 9.5 different consonants word-initially (range 4–16) and 5.7 finally (0–11), out of the 22 consonants possible initially in adult English and 21 finally. The size of the children's inventories in the two word positions was correlated, meaning that children with more differentiated word-initial consonant use were likely to have more different codas in use as well. For comparison, in Vihman et al. (2013), 32 British children (11 of whom were "late talkers") had, on average, inventories of 7.1 consonants (range 5–11), based on 25-word samples recorded at the end of the single-word period, at ages ranging from 15 to 36 months.

    The inventories in Stoel-Gammon's (1987) study typically included the early-learned consonant types word-initially (stops – but not [p] – in all three places of articulation, nasals, [w] and [h]); in addition, the fricatives [f] and [s] occurred in at least half of the samples. In the final position only the voiceless stops, [n], [s], and [r], occurred in half the samples. In a longitudinal overview of the same children's data, Stoel-Gammon (1985) reported that fricatives, affricates, and liquids came into use later than stops and nasals in all positions. No one cluster type occurred in half of the children's inventories in either word position, but the samples showed the very beginnings of cluster use, with a mean of 2.2 different cluster types initially and 1.7 finally.
iii. The mean PCC was 70% (range 43% to 91%). The children with larger inventories also showed greater accuracy. Stoel-Gammon (1987) points out that

accuracy is considerably higher than could be expected, given the relatively small inventories available; thus the "words attempted … contained a disproportionate number of consonants present in [the children's] inventories" (1987: 328). In other words, two-year-olds are continuing to show selection of words to say partially on the basis of their ability to say them.

Return to rhythm: English beyond age two

Although rhythm is an important factor from the earliest period of speech processing, target-like rhythm is not arrived at in production until after some years of language use. Allen and Hawkins (1980) observed that the speech of children acquiring English tends to sound syllable-timed at age one or two, due to the relatively slow speech rate and children's tendency, in the first months of speech production, to give full weight to each syllable and to produce peripheral rather than central vowels, even in unstressed syllables. Allen and Hawkins note the difficulty of assessing the development of phonological rhythm, given the myriad factors that enter into it – the various functions of phonetic duration and the mix of phonological, lexical, syntactic, and stylistic constraints on stress. However, the recent advances in measuring rhythm cross-linguistically mentioned above have led to corresponding advances in developmental accounts: Two recent studies compare children acquiring British English with children acquiring languages with more syllable-timed characteristics.

Mok (2011, 2013) investigated five children each acquiring English and Cantonese as monolinguals (as well as a group of bilingual children) at ages 3 and 2;6 years, respectively. The younger monolingual English children already had significantly more variability in overall utterance duration, more variability in successive syllables, and a lower proportion of vocalic intervals than the monolingual Cantonese, based on recordings of spontaneous speech; the differences in successive syllables reached significance only at age 3. As suggested in our opening discussion of rhythm, the impression of English as stress-timed depends in part on syllable structure. At 2;6 the simpler syllable types CV and CVC accounted for 71% of the syllables produced. Altogether, the five English-speaking children produced a mean of 6 syllables with clusters in any position (and attempted 10 such syllables); clusters were most commonly attempted and produced in the final position (CVCC). The monolingual English-speaking children also produced longer stressed than unstressed syllables in utterance-medial trochaic words.

Payne et al. (2012) investigated the speech of monolingual children acquiring English, Spanish, and Catalan, three each at ages two, four, and six years. They derived measures of the relative proportion of both vocalic and consonantal intervals from acoustic analyses of semi-structured conversations (based on pictured action scenes). They found differences by ambient language, even in the youngest children, with the English-learners producing a lower proportion of vocalic intervals already at age two. However, the variability in consonant intervals, which should be lower in more syllable-timed speech, proved to be higher in the children than in the adults overall and to decrease over time, even in English, despite the

fact that the range of syllable types is greater in adult English than in the two Romance languages and should thus also increase developmentally over time. The reason for this somewhat paradoxical finding is that the relatively simple open syllables of the early years are accompanied by high variability in phonetic consonant duration, due to poor motor control. With increasing age and lexical knowledge the children gain phonetic mastery while at the same time making phonological advances (i.e., increased used of codas and clusters in all word positions), which leads to a more adult-like level of consonantal variability as well as to sharper cross-linguistic differences. Based on both of these studies, then, we can conclude that mastering the rhythmic pattern of English takes at least as long as achieving accurate segmental production.

Conclusion

What then shall we say of English as the language on which to base our understanding of phonological development in general? From the point of view of perceptual processing, English is readily accessible. Its strong lexical stress facilitates segmentation into words (contrast French, with its phonological phrase-based accent, for example) and words are basic to phonological learning.

However, from the point of view of production English seems to be relatively difficult. Although it has more monosyllables than most European languages, a learner advantage, it also has a relatively high proportion of diphthongs, clusters, and, relatedly, syllable types (although clusters are far more common in Slavic languages, for example, and may accordingly be produced more often, if not more accurately, at an early point in lexical development in children learning those languages; see Szreder 2013). We have indicated that stops, nasals, glottals, and glides are produced early in English, as in other languages. We should add that the interdentals, voiced fricatives, and rhotic approximant are typically the last consonants to be acquired. Furthermore, the production of the full range of consonant clusters is seen only after most English consonants have begun to be accurately produced and the characteristic rhythm of English is achieved only by about age six.

In fact, no one language provides an ideal "model" of acquisition:. The starting point is similar for children everywhere, given the biological foundations in prenatal exposure, ancient perceptual capacities, and slow motoric advances, but different ambient languages channel these capacities in different ways even before word use begins to appear.

REFERENCES

Abercrombie, D. 1967. *Elements of General Phonetics*, Edinburgh: University of Edinburgh Press.

Allen, G.D. and Hawkins, S. 1978. The development of phonological rhythm. In: *Syllables and Segments*, A. Bell and

J.B. Hooper (eds.), 173–185, Amsterdam: North-Holland Publishing Company.
Allen, G.D. and Hawkins, S. 1980. Phonological rhythm: definition and development. In: *Child Phonology, 1: Production*, G.H. Yeni-Komshian, J.F. Kavanagh, and C.A. Ferguson (eds.), 227–256, New York: Academic Press.
Anderson, J.L., Morgan, J.L., and White, K.S. 2003. A statistical basis for speech sound discrimination. *Language and Speech* 46: 155–182.
Arvaniti, A. 2012. The usefulness of metrics in the quanitification of speech rhythm. *Journal of Phonetics* 40: 351–373.
Bahrick, L.E., Lickliter, R., and Flom, R. 2004. Intersensory redundancy guides the development of selective attention, perception, and cognition in infancy. *Current Directions in Psychological Science* 13: 99–102.
Bortfeld, H., Morgan, J.L., Golinkoff, R.M., and Rathbun, K. 2005. Mommy and me: familiar names help launch babies into speech stream segmentation. *Psychological Science* 16: 298–304.
Boysson-Bardies, B. de, Hallé, P., Sagart, L., and Durand, C. 1989. A crosslinguistic investigation of vowel formants in babbling. *Journal of Child Language* 16: 1–17.
Boysson-Bardies, B. de and Vihman, M.M. 1991. Adaptation to language: evidence from babbling and first words in four languages. *Language* 67: 297–319.
Braine, M.D.S. 1974. On what might constitute learnable phonology. *Language* 50: 270–299.
Callan, D.E., Kent, R.D., Guenther, F.H., and Vorperian, H.K. 2000. An auditory-feedback-based neural network model of speech production that is robust to the developmental changes in the size and shape of the articulatory system. *Journal of Speech, Language, and Hearing Research* 43: 721–736.
Chomsky, N. 1965. *Aspects of the Theory of Syntax*, Cambridge, MA: MIT Press.
Cooper, R.P. and Aslin, R.N. 1990. Preference for infant-directed speech in the first month after birth. *Child Development* 61: 1584–1595.
Dauer, R.M. 1983. Stress-timing and syllable-timing re-analyzed. *Journal of Phonetics* 1: 51–62.
DePaolis, R.A., Duffy, H., Keren-Portnoy, T., and Vihman, M.M. 2012. IDS differences in naturalistic speech and in stimuli affect the age at which infants succeed in Headturn Preference Procedure tasks. Presented at the International Child Phonology Conference, Minneapolis.
DePaolis, R.A., Keren-Portnoy, T., and Vihman, M.M. 2010. Exploring cultural impact on segmentation and first word recognition. End of Award Report, Economic and Social Research Council.
DePaolis, R., Vihman, M.M., and Keren-Portnoy, T. 2011. Do production patterns influence the processing of speech in prelinguistic infants? *Infant Behavior and Development* 34: 590–601.
DePaolis, R.A., Vihman, M.M., and Keren-Portnoy, T. 2014. When do infants begin recognizing familiar words in sentences? *Journal of Child Language* 41: 226–239.
DePaolis, R., Vihman, M. M., and Nakai, S. 2013. The influence of babbling patterns on the processing of speech. *Infant Behavior and Development* 36: 642–649 .
Engstrand, O., Williams, K., and Lacerda, F. 2003. Does babbling sound native? Listener responses to vocalizations produced by Swedish and American 12- and 18-month olds. *Phonetica* 60: 17–44.
Ferguson, C.A. and Farwell, C.B. 1975. Words and sounds in early language acquisition. *Language* 51: 419–439. Reprinted in Vihman and Keren-Portnoy (2013c: 93–132).
Fernald, A., Taeschner, T., Dunn, J., Papoušek, M., Boysson-Bardies, B. de, and Fukui, I. 1989. A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal

infants. *Journal of Child Language* 16: 477–501.
Gout, A. 2001. Étapes précoces de l'acquisition du lexique [Early steps in lexical acquisition]. Unpublished Dissertation. Écoles des Hautes Études en Sciences Sociales, Paris, France.
Grabe, E. and Low E.L. 2002. Durational variability in speech and the rhythm class hypothesis. In: *Papers in Laboratory Phonology VII*, C. Gussenhoven and N. Warner (eds.), 515–546, Cambridge: Cambridge University Press.
Hallé, P. and Boysson-Bardies, B. de. 1994. Emergence of an early lexicon: infants' recognition of words. *Infant Behavior and Development* 17: 119–129.
Höhle, B. and Weissenborn, J. 2003. German-learning infants' ability to detect unstressed closed-class elements in continuous speech. *Developmental Science* 6: 122–127.
Hunter, M.A. and Ames, E. W. 1988. A multifactor model of infant preferences for novel and familiar stimuli. In: *Advances in Infancy Research*, vol. 5, C. Rovee-Collier and L.P. Lipsitt (eds.), 69–95.
Jaeger, J.J. 1997. How to say "Grandma": the problem of developing phonological representations. *First Language* 17: 1–29.
Jakobson, R. 1949. Les lois phoniques du langage enfantin et leur place dans la phonologie générale. In: *Principes de Phonologie,* N.S. Trubetzkoy (ed.), 367–379. Translated into French by J. Cantineau, Paris: Editions Klincksieck.
Jusczyk, P.W. 1997. *The Discovery of Spoken Language*, Cambridge, MA: MIT Press.
Jusczyk, P.W. and Aslin, R.N. 1995. Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology* 29: 1–23.
Jusczyk, P.W., Cutler, A., and Redanz, N.J. 1993. Infants' preference for the predominant stress patterns of English words. *Child Development* 64: 675–687.
Jusczyk, P.W., Houston, D., and Newsome, M. 1999. The beginnings of word segmentation in English-learning infants. *Cognitive Psychology* 39: 159–207.
Jusczyk, P.W., Friederici, A.D., Wessels, J., Svenkerud, V.Y., and Jusczyk, A.M. 1993. Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language*, 32, 402–420.
Jusczyk, P.W., Luce, P.A., and Charles-Luce, J. 1994. Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language* 33: 630–645.
Kager, R. 1999. *Optimality Theory*, Cambridge: Cambridge University Press.
Kager, R., Pater, J., and Zonneveld, W. (eds.). 2004. *Constraints in Phonological Acquisition*, Cambridge: Cambridge University Press.
Kent, R.D. and Murray, A.D. 1982. Acoustic features of infant vocalic utterances at 3, 6, and 9 months. *Journal of the Acoustic Society of America* 72: 353–363.
Kuijpers, C., Coolen, R., Houston, D., and Cutler, A. 1998. Using the head-turning technique to explore the cross-linguistic performance differences. In: *Advances in Infancy Research*, vol. 12, C. Rovee-Collier, L.P. Lipsitt, and H. Hayne (eds.), 205–220, Stamford, CT: Ablex Publishing Corporation.
Laver, J. 1994. *Principles of Phonetics*, Cambridge: University of Cambridge Press.
Lecanuet, J.-P. 1993. Fetal responsiveness to voice and speech. In: *Developmental Neurocognition: Speech and Face Processing in the First Year of Life*, B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P.F. MacNeilage, and J. Morton (eds.), 237–248, Dordrecht: Kluwer.
Lewkowicz, D.J. 2011. The biological implausibility of the nature–nurture dichotomy and what it means for the study of infancy. *Infancy* 16: 331–367.
Locke, J.L. 1983. *Phonological Acquisition and Change*. New York: Academic Press.
Majorano, M., Vihman, M.M., and DePaolis, R.A. 2014. The relationship between infants' production experience

and their processing of speech. *Language Learning and Development* 10: 179–204.
Mandel, D.R., Jusczyk, P.W., and Pisoni, D.B. 1995. Infants' recognition of the sound patterns of their own names. *Psychological Science* 6, 314–317.
Mattys, S.L., Jusczyk, P.W., Luce, P.A., and Morgan, J.L. 1999. Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology* 38: 465–494.
Maurer, D. and Werker, J.F. 2014. Perceptual narrowing during infancy: a comparison of language and faces. *Developmental Psychology* 56: 154–178.
Maye, J., Werker, J.F., and Gerken, L. 2002. Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* 82: B101–B111.
McCune, L. and Vihman, M.M. 2001. Early phonetic and lexical development. *Journal of Speech, Language and Hearing Research* 44: 670–684.
Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., and Amiel-Tison, C. 1988. A precursor of language acquisition in young infants. *Cognition* 29: 143–178.
Menn, L. 1971. Phonotactic rules in beginning speech: a study in the development of English discourse. *Lingua* 26: 225–251.
Menn, L. 1976. Pattern, control and contrast in beginning speech: a case study in the development of word form and word function. Unpublished PhD thesis, University of Illinois at Urbana-Champaign. Reprinted by the Indiana University Linguistics Club, Bloomington, IN, 1978.
Menn, L. and Vihman, M.M. 2011. Features in child phonology: inherent, emergent, or artefacts of analysis? In: *Where Do Phonological Features Come From? Cognitive, Physical and Developmental Bases of Distinctive Speech Categories*, N. Clements and R. Ridouane (eds.), 261–301, Amsterdam: John Benjamins.
Mok, P.K. 2011. The acquisition of speech rhythm by three-year-old bilingual and monolingual children: Cantonese and English. *Bilingualism: Language and Cognition* 14: 458–472.
Mok, P. 2013. Speech rhythm of monolingual and bilingual children at 2;6: Cantonese and English. *Bilingualism: Language and Cognition* 16: 693–703. doi:10.1017/S1366728912000636.
Nazzi, T., Bertoncini, J., and Mehler, J. 1998. Language discrimination in newborns: toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance* 24: 756–766.
Nazzi, T., Iakimova, G., Bertoncini, J., Frédonie, S., and Alcantara, C. 2006. Early segmentation of fluent speech by infants acquiring French. *Journal of Memory and Language* 54: 283–299.
Oller, D.K. 2000. *The Emergence of the Speech Capacity*, Mahwah, NJ: Lawrence Erlbaum.
Oller, D.K., Wieman, L.A., Doyle, W.J., and Ross, C. 1976. Infant babbling and speech. *Journal of Child Language* 3: 1–11.
Payne, E., Post, B., Astruc, L., Prieto, P., and Vanrell, M.M. (2012). Measuring child rhythm. *Language and Speech* 55: 203–229.
Pike, K. 1945. *The Intonation of American English*, Ann Arbor, MI: University of Michigan Press.
Polka, L. and Sundara, M. 2012. Word segmentation in monolingual infants acquiring Canadian English and Canadian French: native language, cross-dialect, and cross-language comparisons. *Infancy* 17: 198–232.
Priestly, T.M.S. 1977. One idiosyncratic strategy in the aquisition of phonology. *Journal of Child Language* 4: 45–66. Reprinted in Vihman and Keren-Portnoy (2013c: 217–237).
Ramus, F. 2002. Language discrimination by newborns: teasing apart phonotactic, rhythmic, and intonational cues. *Annual Review of Language Acquisition* 2: 85–115.
Ramus, F., Nespor, M., and Mehler, J. 1999. Correlates of linguistic rhythm in the speech signal. *Cognition* 73: 265–292.

Sander, E. 1972. When are speech sounds learned? *Journal of Speech and Hearing Disorders* 37: 55–63.

Shriberg, L. and Kwiatkowski, J. 1982. Phonological disorders III: a procedure for assessing severity of involvement. *Journal of Speech and Hearing Disorders* 47: 256–270.

Smith, N.V. 1973. *The Acquisition of Phonology: A Case Study*, Cambridge: Cambridge University Press.

Stoel-Gammon, C. 1985. Phonetic inventories, 15–24 months: a longitudinal study. *Journal of Speech and Hearing Research* 28: 505–512.

Stoel-Gammon, C. 1987. Phonological skills of two-year-olds. *Language, Speech, and Hearing Services in Schools* 18: 323–329.

Stoel-Gammon, C. and Dunn, J. 1985. *Normal and Disordered Phonology in Children*, Baltimore: University Park Press.

Swingley, D. 2005. Eleven-month-olds' knowledge of how familiar words sound. *Developmental Science* 8: 432–443.

Szreder, M. 2013. The acquisition of consonant clusters in Polish: a case study. In: *The Emergence of Phonology: Whole-word Approaches, Cross-linguistic Evidence*, M.M. Vihman and T. Keren-Portnoy (eds.), 343–361, Cambridge: Cambridge University Press.

Velten, H.V. 1943. The growth of phonemic and lexical patterns in infant language. *Language* 19: 231–292. Reprinted in *Child Language: A Book of Readings*, A. Bar-Adon and W.F. Leopold (eds.), 82–91, 1971, Englewood Cliffs, NJ: Prentice-Hall.

Vihman, M.M. 1993. Vocal motor schemes, variation, and the production-perception link. *Journal of Phonetics* 21: 161–169.

Vihman, M.M. 1996. *Phonological Development: The Origins of Language in the Child*, Oxford: Basil Blackwell.

Vihman, M.M. 2014. *Phonological Development: The First Two Years*, 2nd edition, Oxford: Wiley-Blackwell.

Vihman, M.M. (in press). Perception and production in phonological development. In: *Handbook of Language Emergence*, B. MacWhinney and W. O'Grady (eds.), Oxford: Wiley-Blackwell.

Vihman, M.M. and Croft, W. 2007. Phonological development: toward a "radical" templatic phonology. *Linguistics* 45: 683–725. Reprinted in Vihman and Keren-Portnoy (2013c: 17–57).

Vihman, M.M., DePaolis, R.A., and Keren-Portnoy, T. (2014). The role of production in infant word learning. *Language Learning* 64, Supplement 2: 121–140.

Vihman, M.M., Ferguson, C.A., and Elbert, M. 1986. Phonological development from babbling to speech: common tendencies and individual differences. *Applied Psycholinguistics* 7: 3–40.

Vihman, M.M., Kay, E., Boysson-Bardies, B. de, Durand, C., and Sundberg, U. 1994. External sources of individual differences? A cross-linguistic analysis of the phonetics of mothers' speech to one-year-old children. *Developmental Psychology* 30: 652–663.

Vihman, M.M. and Keren-Portnoy, T. 2013a. Effects of input on early word learning. End of award report, Economic and Social Research Council.

Vihman, M.M. and Keren-Portnoy, T. 2013b. Introduction. In: *The Emergence of Phonology: Whole-word Approaches, Cross-linguistic Evidence*, M.M. Vihman and T. Keren-Portnoy (eds.), 1–14, Cambridge: Cambridge University Press.

Vihman, M.M. and Keren-Portnoy, T. (eds.) 2013c. *The Emergence of Phonology: Whole-word Approaches, Cross-linguistic evidence*, Cambridge: Cambridge University Press.

Vihman, M.M., Keren-Portnoy, T., Whitaker, C., Bidgood, A., and McGillion, M. 2013. Late talking toddlers: relating early phonological development

to later language advance. *York Papers in Linguistics* 2: 47–68.
Vihman, M.M. and Kunnari, S. 2006. The sources of phonological knowledge: a cross-linguistic perspective. *Recherches Linguistiques de Vincennes* 35: 133–164.
Vihman, M.M., Macken, M.A., Miller, R., Simmons, H., and Miller, J. 1985. From babbling to speech: a reassessment of the continuity issue. *Language* 61: 395–443.
Vihman, M.M. and Majorano, M. 2014. The role of geminates in infants' early words and word-form recognition. Presented at the British Association of Academic Phoneticians, April.
Vihman, M.M., Nakai, S., DePaolis, R.A., and Hallé, P. 2004. The role of accentual pattern in early lexical representation. *Journal of Memory and Language* 50: 336–353.
Vihman, M.M., Thierry, G., Lum, J., Keren-Portnoy, T., and Martin, P. 2007. Onset of word form recognition in English, Welsh and English-Welsh bilingual infants. *Applied Psycholinguistics* 28: 475–493.
Vihman, M.M. and Velleman, S. 1989. Phonological reorganization: a case study. *Language and Speech* 32: 149–170. Reprinted in Vihman and Keren-Portnoy (2013: 238–258).
Vihman, M.M., Velleman, S.L., and McCune, L. 1994. How abstract is child phonology? Towards an integration of linguistic and psychological approaches. In: *First and Second Language Phonology*, M. Yavas (ed.), 9–44, San Diego: Singular Publishing. Reprinted in Vihman and Keren-Portnoy (2013c: 259–290).
Vihman, M.M. and Wauquier, S. (in press). Templates in child language. In: *Sources of Variation in First Language Acquisition: Languages, Contexts, and Learners*, M. Hickmann, E. Veneziano, and H. Jisa (eds.), Trends in Language Acquisition Research (TiLAR), Amsterdam: John Benjamins.
Waterson, N. 1971. Child phonology: a prosodic view. *Journal of Linguistics* 7: 179–211. Reprinted in Vihman and Keren-Portnoy (2013: 61–92).
Werker, J.F. and Tees, R.C. (1984). Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development* 7: 49–63.
Whalen, D.H., Levitt, A.G., and Wang, Q. 1991. Intonational differences between the reduplicative babbling of French- and English-learning infants. *Journal of Child Language* 18: 501–516.
White, L. and Mattys, S.L. 2007. Calibrating rhythm: first language and second language studies. *Journal of Phonetics* 35: 501–522.

Appendix

First words of monolingual children acquiring English

Voiceless symbols indicate stops perceived as having short-lag VOT, voiced symbols, pre-voicing, a raised [ʰ], long lag.

Alice (Vihman, Velleman, McCune 1994): **American English, 9–10 months**

[beɪbi]	*baby*	[pɛpɛː], [tɛɪtiː]
[dædi]	*daddy*	[dæ]
[hɑɪ]	*hi*	[hɑːiː], [ʔɑːjɛ], [hɑɪje] [haɪjʌ]…
[mɑmi]	*mommy*	[mːɑnːə]
[noʊ]	*no*	[njæ̃]

Daniel (Danny) (Menn 1971): **American English, 20 months**

[gʊbɑɪ], [bɑɪbɑɪ]	*goodbye, byebye*	[bæbæ, baba, gægæ]
[hɛloʊ]	*hello*	[hwoʊ]
[hɑɪ]	*hi*	[hæ, haɪ]
[noʊ]	*no*	[ono, no, nu]
[noʊz]	*nose*	[o]
[skwəɹl̩]	*squirrel*	[gæ, goʊ]

Daniel (Stoel-Gammon and Dunn 1985): **American English, 12 months**

[bənænæ]	*banana*	[nænæ]
[lɑɪt]	*light*	[ai], [dai]
[ʔʌʔoʊ]	*uh-oh*	[ʔʌʔo]
[wəsðæt]	*what's that*	[wəsæ]

Deborah (Vihman, in press): **American English, 10 months**

[bæ:]	*baa*	[bæ:]
[beɪbi]	*baby*	[be], [pipe], [bebe]
[hɑɪ], [hɑɪjʌ]	*hi, hiya*	[hai], [ai], [haie], [aie], [e:], [a:]
[mʌŋki]	*monkey*	[mam:ɛ]
[ʌʔ:oʊ]	*uh-oh*	[ʌʔ:ɛ]

Emily (Vihman, in press): **American English, 13 months**

[bæ:bæ:], [bɑʊwɑʊ]	*baa, bow-wow*	[pæpæ], [bæbæ], [ʔapɪæ], [pæ:]
[bi:dz]	*beads*	[bi], [p^{h}i]
[dædi]	*daddy*	[tæ], [hadatɛ]
[ʌp]	*up*	[ʌp], [ʌpə], [ʌpije], [æb]

Jacob (Menn 1976): **American English, 13 months**

[noʊ]	*no*	[nʌ:::], [ŋɛʌ]
[dʒeɪkəb]	*Jacob*	[dikʌ], [dɛikʌ], [gɛikʌ], [æku], [dɛikʌ], [æku]
[θæŋkju:]	*thankyou*	[didʌ], [dɪdɐjdi], [łɛjʌ], [da^{ˈ}za], [di], [da^{ˈ}dʌ], [bɛ], [dɜt], [gɑdu], etc.
[ðɛɹ]	*there*	[dɑ], [dʌm], [dʌh], [dɛ], [dæ]
[t^{h}oʊst]	*toast*	[dœʌ]

Joan (Velten 1943): **American English, 11–12 months**

[bæŋ], [bɑɾəl]	*bang, bottle*	[ba]
[bʌs], [bɑks]	*bus, box*	[bas]
[pʊɾɔn]	*put on*	[baza], [ba:za]
[ðæt]	*that*	[za]
[ʌp]	*up*	[ap]

Jonah (Vihman 1996: App. B): **American English, 13 months**

[bɑɾəl]	*bottle*	[bwɪdʊ]
[bɑʊwɑʊ]	*bow(wow)*	[bɑʔ], [bʊɑ], [bæ]
[ɛdgəɹ]	*Edgar (dog's name)*	[dɑdɑ]
[noʊ]	*no*	[ənæ::]

Jonathan (Braine 1974): **American English, 15 months**

[hɑɪ]	*hi*	[ʔai]
[dʒus]	*juice*	[du]
[noʊ]	*no*	[do]
[si:]	*see*	[di]
[ðæt], [ðɛɹ]	*that, there*	[dæ, dʌ, da, dɛ]

Leslie (Ferguson, Peizer, and Weeks 1973): **American English, 11 months**

[dædi]	*daddy*	[dædæ]
[dɔgi]	*doggie*	[gaga]
[mami]	*mommy*	[mama]
[pæti], [pædi]	*patty(-cake)*	[bæbæ]

Molly (Vihman and Velleman 1989): **American English, 10–11 months**

[beɪbi]	*baby*	[bæpæ]
[kˌjækəɹ]	*cracker*	[pɑkæ], [kwɑ], [wæhk], [pækwɑ], [kʌk]
[mu::]	*moo*	[meʔje]
[nɑɪʔnɑɪt]	*night-night*	[hʌn:ʌ], [noʊnæ]

Sarah (Stoel-Gammon and Dunn 1985): **American English, 11 months**

[beɪbi]	*baby*	[bebi]
[bɑɪbɑi]	*byebye*	[baɪbaɪ]
[dɑgi]	*doggie*	[dɔgi]
[dʒu:s]	*juice*	[dus]
[mɑmɑ]	*mama*	[mama]

Sean (Vihman and Kunnari 2006): **American English, 12 months**

[Algçn]	*allgone*	[ɔdæ:]
[bu:]	*boo*	[pʊ]
[dɑg]	*dog*	[tak]
[tɪk]	*tick*	[tɛʰ], [tɪʔ], [tˌj], [tɔ̊t]
[wʊf]	*woof*	[wʊ], [ʔʊʔ], [ʔoʊ]
[skwəɹəl]	*squirrel*	[gæ, goʊ]

T. (Ferguson and Farwell 1975): **American English, 11 months**

[dædi]	*daddy*	[dæji, dæi]
[dɑg]	*dog*	[dɔ]

[hɑɪ]	*hi*	[ai], [hai]
[si:]	*see*	[hi]

Timmy (Vihman, Velleman, and McCune 1994): **American English, 11 months**

[bɔl]	*ball*	[pæ], [bæ], [ʔəpæ], [ab:a]
[blɑk]	*block*	[əpʰə], [ʔʌβæ], [pæ]
[kʰɑɹ]	*car*	[kɑə], [ɑk:ɑʰ]
[kʰɪɾi]	*kitty*	[khə̰], [kʰɑ̰], [kɑkɑ], [ʔukɑ]…
[kʰwæʔkʰwæk]	*quack-quack*	[kʰə̰], [kʰɑ], [kʰɑʰkɑʰ], [gaga]…

Tomos (Vihman, in press b): **UK English, 17 months**

[bædʒə]	*Badger*	[babm̩:], [bʌbm̩]
[bæŋ]	*bang*	[ba], [bæ], [baʊ], [da]
[haɪja]	*hiya*	[jaja], [dajæ:]
[nəʊ]	*no*	[na], [næ], [nə]
[ta]	*ta 'thank you'*	[ba], [pa], [ba:], [ɹa:]

Will (Stoel-Gammon and Dunn 1985): **American English, 12 months**

[ɑldʌn]	*all done*	[dada], [ada]
[daʊn]	*down*	[dæ], [dʌ], [dau]
[lɑɪt]	*light*	[di]
[ʃu:z]	*shoes*	[tsis, θiz]
[ʌʔ:oʊ]	*uh-oh*	[ʔʌʔo], [hʌho]

20 Variables Affecting L2 Pronunciation Development

PAVEL TROFIMOVICH, SARA KENNEDY, AND JENNIFER ANN FOOTE

Introduction

John Austin, a renowned British philosopher of language, once noted in one of his lectures on language that "the uttering of the sentence is the doing of an action, which […] would not *normally* be described […] as 'just' saying something" (Austin 1975: 5, original emphasis). Central to this statement is the idea that language is not simply a cognitive system of mental representations and rules (e.g., Chomsky 1957), but rather a tool used by individuals to accomplish real goals or actions through interaction. Along with Grice and Searle, other fellow philosophers of language, Austin laid the foundations for what later Herbert Clark (1992) termed a "language-as-action" tradition in linguistics. The view of language as action, according to Clark, presupposes that language is used by *participants* (real people who often have defined roles, such as a test-taker, customer, or employer) in order to accomplish certain interactive *social processes* (real-world goals, such as completing a business transaction, making a case in court) as part of *collective actions* (contextualized instances of language use).

The language-as-action view provides a fitting framework for discussing pronunciation. Pronunciation lies at the core of oral language expression, identifying individual speakers and speaker communities. Pronunciation is also central to language use in social, interactive contexts because pronunciation embodies the way that the speaker and the hearer work together to establish and maintain common ground for producing and understanding each other's utterances. Last but not least, pronunciation, as a way of speaking, is intimately linked to the particular places, times, and situations of language use, such that, for instance, giving a public lecture at a university, discussing a hockey match at a bar, or sharing bad news with a loved one in a hospital would involve different ways of speaking.

The Handbook of English Pronunciation, First Edition. Edited by Marnie Reed and John M. Levis.

In this chapter, we adopt Clark's (1992) view of language as action to discuss several influences on second language (L2) pronunciation development (variables). Such influences are categorized according to the three properties thought to describe language use: participants, social processes, and collective actions. We synthesize current scholarship and scholarly debates with respect to each factor described and then outline possible avenues for future research. We conclude by discussing some viable theoretical views of L2 pronunciation development, particularly those that embrace the multidimensional nature of pronunciation learning. Our overall objective is to show that L2 pronunciation learning, a challenging and exciting area for researchers, teachers, and learners alike, is a complex and multidimensional phenomenon.

Participants

The first cluster of variables we discuss in relation to L2 pronunciation development fits well within the "participants" category described by Clark (1992). According to Clark, language use is quintessentially a human activity, with individuals often having defined roles in various episodes of language use (e.g., teacher, student, employer, bystander, etc.). As participants involved in language use, individuals therefore bring to the interaction a number of person-specific factors, or variables that reflect in some way their individual differences or capacities (e.g., age, aptitude, perception ability). Several of these person-specific variables or individual differences that can impact L2 pronunciation development are discussed here.

Age

One of the most widely discussed (and hotly debated) factors in relation to L2 pronunciation development is learners' age. The idea that L2 learning (and learning L2 pronunciation in particular) might depend on learners' age dates back to the writings of Penfield and Roberts (1959). These researchers were among the first to propose that in order for a child to learn a language to native-like mastery, exposure to that language must occur within a certain developmental "window" described as a critical or a sensitive period. This idea was later taken up by Lenneberg (1967), who speculated that a critical period for language, which was biologically determined through brain maturation, ended around the age of puberty. The critical/sensitive period for language learning, of the kind proposed by Lenneberg, thus involves a certain biologically determined period of sensitivity to language followed by a decline in the capacity to learn it (see Bornstein 1987 for more on critical/sensitive periods).

To date, researchers have gathered an extensive body of evidence supporting the basic assumption underlying the notion of a critical/sensitive period – that learning an L2 beyond early childhood appears to result in often incomplete, nonnativelike mastery of the language. With respect to L2 pronunciation, for example, there is ample evidence that children, while often initially slower at L2 learning,

eventually outperform adults on a variety of tasks, and that even the most successful adult learners are seldom fully native-like in their L2 pronunciation (e.g., Abrahamsson and Hyltenstam 2009; Aoyama et al. 2008; Bongaerts et al. 1997; Flege, Yeni-Komshian, and Liu 1999).

At the heart of the sensitive period controversy is whether L2 "age effects" are determined by a biologically-driven critical/sensitive period or instead arise as a consequence of other factors. Some researchers, like Lenneberg, support the notion of a biologically-determined critical/sensitive period for L2 learning. For example, Pulvermüller and Schumann (1994) attribute older children's and adults' diminishing ability to learn an L2 to a gradual decline in neuronal plasticity in specific areas of the brain (see also Jacobs 1988). Others, however, refute the existence of a biologically-determined critical/sensitive period, instead linking age effects to a variety of social-educational factors (e.g., Jia and Aaronson 2003; Flege, Yeni-Komshian, and Liu 1999; Moyer 1999) or cognitive variables (e.g., Hakuta, Bialystok, and Wiley 2003). Still others hypothesize that age effects do not necessarily reflect age-bound neurobiological limitations alone but arise as a consequence of the act of prior learning itself, such that speech perception and production become specialized for the processing of native language (L1) input (Baker et al. 2008; McCandliss et al. 2002).

An examination of the literature on child–adult differences reveals a number of plausible interpretations of L2 age effects, including those with neurobiological, linguistic, social, attitudinal, experiential, and cognitive underpinnings (see Birdsong 2009 and DeKeyser 2012). In our view, one of the most promising (and empirically testable) interpretations for age effects in L2 pronunciation learning relates to differential involvement of memory systems in child versus adult L2 pronunciation learning (DeKeyser 2012; Paradis 2009; Ullman 2005). The two memory systems in question are declarative memory, responsible for the learning of form-meaning relationships stored in the lexicon, and procedural memory, responsible for the learning of grammar and pronunciation. The information stored in declarative memory is generally explicit (open to conscious awareness), whereas procedural memory is responsible for implicit learning (learning without awareness). For instance, it has been proposed that adolescent and adult L2 learners mostly rely on analytical, declarative, explicit learning mechanisms in learning aspects of L2 morphosyntax, whereas children have access to procedural, implicit learning mechanisms (Abrahamsson and Hyltenstam 2008; DeKeyser, Alfi-Shabtay, and Ravid 2010). Recent evidence from L2 pronunciation research is compatible with this interpretation (Archila-Suerte et al. 2012; Saito 2013). In fact, based on this and similar evidence, DeKeyser (2012) recently suggested that the core question to guide current and future research on L2 age effects is "whether there is a specific period of decline in the ability for implicit language learning" (2012: 446). With respect to L2 pronunciation development, therefore, it remains to be shown whether and to what extent adult learners rely on implicit learning and whether pronunciation teaching activities could harness implicit learning strategies (for preliminary evidence, see Trofimovich, McDonough, and Neumann, 2013; Trofimovich, McDonough, and Foote, 2014).

Cross-language perceptual similarity

One of the most salient influences on L2 pronunciation development can be traced to learners' L1. It is a common observation that L2 learners' perception errors and foreign accents are in large part specific to their L1. In previous research, L1-based influences on L2 pronunciation development have been studied through typological L1–L2 comparisons, for example, by comparing the status of particular phonemic and phonetic categories cross-linguistically (for a review, see Davidson 2011). However, L1 effects on L2 pronunciation arguably also reflect an *individual* learner's ability to perceive similarities or differences across the two languages. The assumption here is that L1 influences on L2 pronunciation are ultimately a matter of perception, or the degree to which aspects of L2 pronunciation (e.g., individual segments or prosodic patterns) are filtered through and recognized in terms of the learners' L1 (Strange 2007).

From this vantage point, one way to characterize L1 effects on L2 pronunciation is through the construct of cross-language perceptual similarity. Cross-language similarity refers to how perceptually similar or dissimilar listeners treat specific aspects of pronunciation in their L1 and L2. There is evidence that the degree of perceived similarity (or dissimilarity) between L1 and L2 sounds may determine how L2 sounds are perceived and produced (Baker and Trofimovich 2005; Guion et al. 2000; Strange et al. 2011). For example, Japanese learners of L2 English may perceive and produce English /r/ more accurately than English /l/ (Flege, Takagi, and Mann 1995) because they are more likely to perceptually differentiate English /r/, but not /l/, from Japanese /r/ (Aoyama et al. 2004). In this situation, cross-language dissimilarity renders one L2 segment (English /r/ in this case) easier to learn than another (English /l/).

As the above example suggests, L2 perception and production appear to depend on the perceived distance between L1 and L2, such that (depending on the particular relationship between L1 and L2) cross-language similarity can either help or hinder L2 perception and production. This idea has been central to two influential models of L2 speech learning – the Perceptual Assimilation Model (Best and Tyler 2007) and the Speech Learning Model (Flege 2002). Both models hold that perception and production of specific aspects of L2 pronunciation depend on L2 learners' ability to detect cross-language differences at the level of pronunciation. Both models also assume that only a perceptual measure of cross-language similarity – as opposed to those based on comparisons of acoustic properties, sound categories, prosodic units, or distinctive features – qualifies as a direct and predictive measure of L2 perception and production difficulty (Strange 2007). Directly estimating cross-language perceptual similarity, for example, involves having L2 learners compare target L2 vowels to vowels in the learners' L1 using perceptual identification or similarity rating tasks (e.g., Strange et al. 2011).

In the past two decades, explorations of cross-language perceptual similarity, usually carried out within the conceptual framework of either or both of the above models, have received careful attention in L2 pronunciation research, with the specific aim of determining the perceptual difficulty and learnability of different L2

pronunciation features. For example, in an early study, Bohn and Flege (1990) demonstrated that the perceptual relationship between English /æ/ and German vowels, established in a cross-language perceptual identification experiment, explained native German listeners' difficulty in discriminating the English /æ/-/ɛ/ contrast (as in *bat-bet*). In a seminal study a decade later, Guion et al. (2000) argued that the perceptual similarity between English and Japanese consonants both explained and predicted which English consonants are most difficult for Japanese learners of English to perceive. The encouraging outcomes of this research motivate further investigations of cross-language similarity to determine the difficulty and learnability of different aspects of L2 pronunciation at different stages of learning. Productive future avenues of research might involve investigations of the role of cross-language similarity in the learning of prosodic, as opposed to segmental, aspects of L2 pronunciation and comparisons of cross-language similarity for learners of different ages, with a view to explaining child–adult differences in L2 pronunciation development (Baker et al. 2008). Future research should also explore pedagogical uses of cross-language perceptual similarity, as part of cross-language awareness building activities and perceptual training (see Thomson 2012).

Aptitude

Language aptitude refers to a cluster of cognitive variables believed to underlie the human capacity for language learning. Although the precise variables considered as part of language aptitude are often specific to particular instruments used to measure it (Carrol and Sapon 1959; Grigorenko, Sternberg, and Ehrman 2000; Pimsleur 1966; Sparks et al. 2011), language aptitude commonly encompasses aspects of short-term memory, phonetic coding (ability to encode and retain auditory sequences), grammatical sensitivity (ability to recognize grammatical functions of words), rote learning (ability to form sound-meaning associations), imitation or mimicry, inductive learning (ability to infer rules or patterns from linguistic information), musical ability, as well as transfer and combination skills (ability to apply inferred patterns to new contexts and to synthesize information). Despite decades of productive research on language aptitude (Skehan 2012), there has been little systematic research on the relationship between various subcomponents of language aptitude and L2 pronunciation learning. Most of the research carried out within the aptitude tradition has examined the contribution of musical ability to the learning of L2 pronunciation, testing the basic assumption that there is an association between musical ability and the quality of L2 pronunciation.

Musical ability typically refers to an individual's ability to "hear" (internalize) music that is no longer present in the physical environment, a skill that Gordon (1995) termed "audiation". For example, upon hearing two musical phrases played consecutively, listeners with greater musical ability, as compared to listeners with weaker musical ability, would presumably be able to judge whether the two phrases are similar in their melodic contour (overall pattern of pitch rises and falls), even if the two phrases differed in the overall number of notes. Musical ability, defined in this manner, is often measured using standardized tests, which

target several aspects of this ability, including pitch, intensity, rhythm, timbre, tonal memory, and timing (Bentley 1966; Gordon 1995; Seashore 1919; Wing 1968).

Although L1 research has shown an important association between musical ability and speech processing, particularly with respect to the music–prosody links (Palmer and Hutchins 2006), the relationship between musical ability and L2 pronunciation remains unclear. Some researchers who have investigated this relationship reported a positive correlation between musical ability and L2 pronunciation (Arellano and Draper 1972; Milovanov et al. 2010; Slevc and Miyake 2006). However, many others have failed to reveal any clear relationship between these two variables (Dexter and Omwake 1934; Flege, Munro, and MacKay 1995; Pimsleur, Stockwell, and Comrey 1962; Tahta, Wood, and Loewenthal 1981). The link between musical ability and L2 speech perception has been even more elusive, essentially because this relationship has been studied much less extensively. For example, Slevc and Miyake (2006) showed that a standardized measure of musical ability accounted for up to 12% of variance in native Japanese speakers' perception of L2 (English) contrasts in words, sentences, and spoken texts (see also Milovanov et al. 2008; Pimsleur, Stockwell, and Comrey 1962). However, in several other studies, no association between musical ability and L2 perception was found (Arellano and Draper 1972; Milovanov et al. 2010). Clearly, more research is needed to enhance our theoretical understanding of the link between musical ability and L2 pronunciation. At the practical level, it would also be important to determine how L2 pronunciation teaching could be made more effective through the use of music-based activities, particularly for the teaching and learning of L2 prosody.

Social processes

The second cluster of variables we discuss in relation to L2 pronunciation development falls under the category of "social processes" identified by Clark (1992). For Clark, the primary goal of language use resides not simply in the act of speaking, but rather in accomplishing a given social goal (e.g., expressing an opinion or getting someone to do something). From this vantage point, then, language learning cannot be considered outside its contexts, which implies that a number of social variables (e.g., ethnicity, motivation) can have a measurable impact on language learning, including L2 pronunciation development. In this section, two of these variables are discussed.

Motivation

Although motivation can be understood as a purely cognitive phenomenon (a subcomponent of language learning aptitude or an individual difference factor), recent research on motivation has firmly placed this variable within the realm of socially situated learning (for a recent review, see Ushioda and Dörnyei 2012). Broadly speaking, motivation refers to a cluster of variables dealing with the willingness, interest, and desire of the language learner to engage in a learning process.

The construct of motivation is tightly linked to its measures, which in the context of L2 pronunciation research have been operationalized as scalar ratings in response to simple statements, such as *English is important for success at work/school* (Flege, Yeni-Komshian, and Liu 1999), or participants' responses to open-ended questions, such as *What is your motivation for studying German at this time?* (Moyer 1999).

Motivation has long been believed to influence L2 pronunciation development (Guiora, Brannon, and Dull 1972; Oyama 1976), yet its precise role has been elusive. For example, in a study of Italian-born immigrants to the United States, Oyama (1976) found no evidence of a relationship between the participants' L2 accent scores and their self-rated motivation while learning or their self-rated motivation to improve their English. Similarly, Flege and his colleagues showed that measures of motivation had a minimal contribution (accounting for less than 2% of total variance) to measures of L2 accent in large samples of native Korean and Italian learners of L2 English in the United States (Flege, Munro, and MacKay 1995; Flege, Yeni-Komshian, and Liu 1999). In contrast, Moyer (1999), who studied highly proficient L2 German speakers enrolled in a German university-level program, found a significant overlap between the speakers' professional motivation (i.e., the importance of German for their future professional lives) and the nativeness of these speakers' speech (see also Bongaerts et al. 1997).

Perhaps one reason for the inconclusive findings thus far is that motivation has rarely been explored in depth in relation to L2 pronunciation development, but has instead been typically treated as a moderator variable measured using a few simple statements. As a rare exception, Polat (2011) recently reported a significant relationship between L2 accent scores and introjection (engaging in a learning activity because of self-imposed sanctions, for instance, to avoid guilt) and integration (engaging in a learning activity for reasons of self-enjoyment and self-fulfillment) for a large sample of young Kurdish learners of Turkish. Both introjection and integration were among several motivational orientations studied within Deci and Ryan's (1985) self-determination theory of motivation, revealing a complex interaction between motivational orientation, L2 accent, and speakers' gender. Future research on L2 pronunciation development would benefit from similar detailed investigations of motivation in L2 pronunciation learning, especially those carried out within the L2 motivational self-system (Dörnyei and Ushioda 2009) and the Willingness to Communicate framework (MacIntyre et al. 1998), as well as those featuring in-depth qualitative measures of motivation.

Ethnic and personal identity

In order to learn an L2, individuals or groups of individuals come into contact with other individuals or groups, increasing the chances that matters of personal and group identity become salient. Ethnic identity can be broadly defined as a subjective experience of being a part of an ethnic group (Ashmore, Deaux, and McLaughlin-Volpe 2004), and in the case of L2 learning, the ethnic groups in question are learners' own (ancestral) ethnic group and the target language (L2) community. There is relatively little research documenting how L2 learners'

identification with their own ethnic group and with the L2 community impacts L2 pronunciation. At least hypothetically, learners may refrain (overtly or covertly) from acquiring an L2, especially if they fear that the vitality of their ethnic group is threatened (Taylor, Meynard, and Rheault 1977). This situation may reflect subtractive bilingualism or assimilation (Giles, Bourhis, and Taylor 1977), whereby individuals (usually members of a minority group) acquire the language of a majority group and often lose their own language and culture. Alternatively, language learners may embrace L2 learning despite a strong sense of ethnic group identity (Ellinger 2000). For example, 95% of the 100 multicultural students surveyed by Derwing (2003) in Alberta indicated their desire to pronounce English like a native speaker and felt that their sense of personal identity was not threatened. This situation illustrates additive bilingualism or integration (Giles, Bourhis, and Taylor 1977), whereby individuals add a new language and culture without losing their own.

The relationship between ethnic identity and L2 pronunciation learning has recently been studied by Gatbonton and her colleagues. These researchers explored whether several aspects of the ethnic identity construct (e.g., strength of identification with one's ethnic group, support for the group's sociopolitical aspirations) are related to measures of L2 pronunciation. For native French speakers of L2 English in Quebec, where French and English are respectively majority and minority languages, both positive and negative identity-pronunciation links were found. Those speakers who expressed stronger political views (e.g., support for Quebec's independence from Canada) were judged as being more accented, less comprehensible, less fluent, and less proficient overall in their L2 English. However, the speakers who had a double-positive orientation (i.e., a positive orientation towards their own ethnic group and the L2 community) were also those who were considered by native listeners to be most proficient in English (Gatbonton and Trofimovich 2008). For Latvian and Russian bilingual speakers in Latvia, where Latvian is the majority and Russian is a minority language, identity-pronunciation links depended on the group studied. For Latvians, a strong sense of ethnic identity was related negatively to their self-rated L2 (Russian) ability. In contrast, for Russians, no such negative associations emerged for their L2 (Latvian), suggesting that these speakers may have preferred (overtly or covertly) not to associate their strong ethnic beliefs with the ability to speak the majority language, perhaps in order to both maintain a strong sense of identity and also to gain access to the social and economic benefits associated with speaking Latvian (Trofimovich, Turuševa, and Gatbonton 2013). These findings show that ethnic groups residing in contact may relate issues of ethnic identity to L2 pronunciation in rather distinct ways, potentially influencing the rate and success of L2 pronunciation learning.

Apart from ethnic identity, several other aspects of the identity construct have been studied in relation to L2 pronunciation development. In a study of nine Americans living in Norway, for instance, Lybeck (2002) showed that the extent of speakers' social and cultural integration, defined as participation in "supportive exchange networks within the target culture" (2002: 184), were related to the accuracy of these speakers' L2 pronunciation (see also Thompson 1991). In another

study, Hansen (1995) examined the relationship between the degree of acculturation and strength of foreign accent for 20 native German immigrants to the United States. She found that acculturation, and especially the degree to which participants engaged in intercultural activities, was negatively associated with L2 accent, such that more acculturated individuals had a more native-like L2 accent (see also Polat and Mahalingappa 2010). Marx (2002) documented a case study of an L2 learner of German, providing a longitudinal perspective on the interplay between L2 accent, on the one hand, and the construction of L2 identity through patterns of language use, on the other. Common to all these studies is the link between L2 pronunciation and identity, most clearly seen through patterns of learners' social, cultural, and linguistic integration.

One broad conclusion that cuts across all strands of identity research is that matters of ethnic identity – construed within the broader sociopolitical setting and a narrower context of a particular learning situation – have consequences for L2 pronunciation development. Indeed, it is plausible that at least some learners would not reach expected levels of L2 proficiency because their language learning needs may clash with their sense of identity. Therefore, a fruitful area of future thinking in this regard would be to consider how language learning motivation, matters of identity (which should include cultural and linguistic patterns of language use), and teaching and learning practices interact to make language learning efficient and enjoyable for L2 learners.

Collective actions

The third cluster of variables we discuss in relation to L2 pronunciation development can be characterized under the broader category of "collective actions", which for Clark (1992) referred to socially coordinated activities performed by more than a single speaker. Aside from a handful of exceptions (e.g., technology-mediated individual practice, self-study), pronunciation learning is inherently an interactive process, taking place in the social context of a language classroom or in a given naturalistic environment (e.g., workplace, community). Therefore, several contextual factors (e.g., pertaining to the quality and quantity of language experience and use) have the potential of influencing L2 pronunciation development. In the following section, several of these factors are discussed.

Amount of experience

When it comes to L2 pronunciation, it is not always the case that the more experience L2 learners have with the language, the better the outcomes of L2 pronunciation learning will be. While some cross-sectional studies have demonstrated that a longer length of residence (LOR) in an L2 environment is associated with more native-like or more favorably rated pronunciation (e.g., Flege, Yeni-Komshian, and Liu 1999; Trofimovich and Baker 2007), this relationship is not always straightforward. Flege and Fletcher (1992) found that the role of LOR in the L2 pronunciation

of L1 Spanish speakers was important only for speakers who had recently arrived in the L2 environment and not for long-time residents. Similarly, a year-long longitudinal study of newly arrived L1 Mandarin and Slavic immigrants to Canada showed significant early improvement in their pronunciation of some but not all English vowels. This improvement slowed considerably after six months of residence (Munro and Derwing 2008).

There is clear and substantial evidence that the number of years an L2 speaker lives in an L2 setting can be less important to L2 pronunciation than other measures of experience. For example, LOR did not predict the accent ratings in English received by Russian immigrants to the United States, whether they had arrived as children or as adults (Thompson 1991). A similar result was found in a longitudinal study of Japanese adult newcomers to the United States, whose production of individual sounds and accentedness ratings did not change after a year of residence (Aoyama et al. 2008). In the same way, Flege et al. (2006) found that Korean adult immigrants to the United States with LORs of three and five years received similar ratings of accentedness in English. Clearly, L2 speakers' LOR may play a minimal role in their development of L2 pronunciation.

Language use

With more detailed measurements of language use, researchers have shown more nuanced relationships between language experience and L2 pronunciation development. Flege, Frieda, and Nozawa (1997) recorded English speech samples from adult speakers who had arrived in Canada from Italy as children. The speakers who reported relatively high daily use of Italian received less native-like accent ratings than the speakers who reported relatively low use. Derwing, Munro, and Thomson (2008) tracked newly arrived L1 Mandarin and Slavic immigrants to Canada for one year and found that only the Slavic speakers showed significant improvement in the ratings received for fluency and comprehensibility. The Slavic group also reported significantly more exposure to English talk radio and significantly more extended interactions with English speakers than did the Mandarin group (see also Derwing and Munro 2013).

The benefits of language use are not restricted only to using the language productively. In a series of studies, Au and his colleagues have shown than children who simply overheard a language spoken around them in early childhood, but never overtly used it for interaction, were judged to be more native-like and less accented years later, compared to "typical" adult L2 learners (Au et al. 2002, 2008). Beneficial effects of listening-only experience on L2 pronunciation development were also reported for children learning an L2 in an instructed classroom setting. The learners who were only exposed to listening and reading input for about one year of instruction performed on a speaking task similarly to learners taught through "traditional" practice that involved speaking (Trofimovich et al. 2009).

A link between language use and L2 pronunciation has been found consistently for L2 speakers in academic settings. Yeni-Komshian, Flege, and Liu (2000) found that L1 Korean university students in the United States who tended to use English

to a greater extent were rated as having more native-like English accents. The opposite relationship was found between students' accentedness ratings and their self-reported percentage use of Korean. Moyer (2011) studied 42 L2 English students at an American university and found significant relationships between accent and various measures of L1 and L2 use. Students reporting a greater number of hours of weekly L1 use were rated as being more accented. In contrast, students reporting more weekly hours speaking the L2, interacting in the L2, and using the L2 with roommates, host families, or native English speaker friends were judged as being less accented.

Relationships between language use and L2 pronunciation have also been found for languages other than English. Díaz-Campos (2004) found that L1 English learners of Spanish who reported using Spanish 4 hours or more per week outside class had more native-like pronunciation than those reporting 0–3 hours per week. Guion, Flege, and Loftin (2000) also showed that Quichua-Spanish bilinguals who reported relatively high use of their L1 (Quichua) were perceived as being more accented in Spanish than bilinguals reporting relatively low L1 use. A more complex relationship between language use and pronunciation was discovered by Yager (1998), who observed language use and measured gains in pronunciation ratings over seven weeks for beginner-, intermediate-, and advanced-level learners of Spanish. No significant relationship was found between language use and pronunciation gains for intermediate-level learners. For beginner-level learners, the more interactive contact in Spanish they reported at the beginning of the seven-week period, the larger the gain in pronunciation ratings at the end of seven weeks. For both beginner- and advanced-level learners, the more noninteractive contact they reported in the first week, the lower the pronunciation gains at the end of the seven-week period. Thus, in both cross-sectional and longitudinal studies, language use has proved to be consistently connected to L2 pronunciation.

Study abroad

The importance of L2 use outside the classroom for the development of L2 speech is a key argument for the existence of study abroad (SA) programs. SA programs allow students who are usually in post-secondary institutions to study in another country for a limited time. SA students living in an L2 environment have many more opportunities to use the L2 across many domains, as opposed to domestic students, who remain in their home country while studying the language. This potential for increased L2 experience is assumed to confer an advantage to SA students in developing their language proficiency and especially their pronunciation.

Recent research on SA students has shown mixed results regarding the learning of L2 pronunciation. Students of Spanish residing in Spain for one semester significantly improved in oral fluency measures, unlike students from the same university program who remained in the United States (Segalowitz and Freed 2004). The same students' pronunciation accuracy for particular sounds varied, depending on the sound; however, there was generally little difference between SA and domestic students in pronunciation gains (Díaz-Campos 2004).

Although learners' experiences with an L2 are thought to be quite different in SA programs as opposed to programs of study in their home countries, each type of program can take several forms. Martinsen et al. (2010) investigated 25 learners of Spanish in three types of settings over seven weeks: a traditional SA program, a service learning SA program, which involved an additional 5–15 hours a week of service benefiting the community, and foreign language housing (a student residence at learners' home university that was explicitly classified as a residence in which only Spanish was used). None of the three groups demonstrated significant improvement in pronunciation ratings received at the beginning and end of the seven-week period. Generally, findings from existing research suggest that although L2 learners' oral fluency may improve after studying in an L2 environment for a short term, their L2 pronunciation may improve only in certain aspects or may be no different from the pronunciation of learners who study in non-SA environments.

In future research on language experience and use, including SA research, researchers need to continue refining methodologies for capturing fine-grained aspects of language use. In-depth qualitative methods of inquiry (e.g., Kurata 2010; Piller 2002) and extensive audio/video observations of speakers' language use in authentic environments (e.g., Lamarre 2013), especially through mobile technology, seem to be promising research tools for helping researchers link aspects of language experience with L2 pronunciation development.

Theoretical frameworks

It is clear from the preceding discussion that a multitude of factors can shape L2 pronunciation development. Therefore, one challenge for L2 researchers is to conceptualize the influence of these and potentially many other variables within coherent and testable theoretical frameworks that link person-specific, social, and experiential factors to L2 learning outcomes. Several existing theoretical proposals are promising in this respect. In the field of cognitive psychology, Dynamic Systems Theory (de Bot, Lowie, and Verspoor 2007) is one such theoretical proposal. The dynamic systems view presupposes that language learning is an iterative (repetitive) process characterized by variability both within and across individuals. This process occurs on many time scales (e.g., within an interaction, across lessons, during semesters of course work, throughout years of language experience) and features a number of developmental stages (called "attractor states"). Van Geert, Steenbeek, and van Dijk (2011) have recently applied this theory to account for socially mediated L2 learning, thus encompassing both cognitive and social aspects of L2 development. In van Geert, Steenbeek, and van Dijk's model, language development occurs through the interaction between a novice (learner) and an expert (teacher), with learning determined by the interplay between the situation-specific goals of the learner (e.g., the need to acquire certain knowledge, to exert less effort in learning, to preserve aspects of own ethnic identity) and the goals of the teacher (e.g., the need to complete certain learning tasks, to motivate

learners, to satisfy requirements from the employer or the curriculum). Learning is thus conceptualized within this model as a continuous, dynamic adaptation of teacher and learner behaviors, with teachers adapting their actions to the perceived needs of learners.

A related theoretical view that appears to be promising for modeling the multidimensional nature of L2 pronunciation development is the sociocognitive approach (Atkinson 2011). This approach is based on the idea that language development is determined by a dynamic interaction between the mind, body, and world. This implies that people's cognitive states, such as person-specific individual variables and mental representations (i.e., the mind), are instantiated in overt behaviors, such as bodily actions, orientations, or emotions (i.e., the body), which are in turn fully embedded in particular social contexts (i.e., the world). As in the language-as-action tradition (Clark 1992), language is seen here as an instrument of social action, as a flexible and adaptable tool of effecting change in a given social environment (e.g., ordering a meal or persuading a listener). Language development is also conceptualized as a gradual, interactive adaptivity or alignment of the learner with a sociocognitive learning environment. For example, a learner might align with the teacher within a given social interaction in a classroom in terms of the complexity of utterances, body gestures, voice volume, and rate of speech (Atkinson et al. 2007; Churchill et al. 2010). This view of learning as social and cognitive alignment, which is compatible with both cognitive research on interactive alignment (Pickering and Garrod 2004) and social psychological research on social accommodation (Giles and Powesland 1975), appears to be very promising for conceptualizing L2 pronunciation development (see Trofimovich 2013 for an initial attempt). The chief benefit is that researchers might use the sociocognitive approach (and Dynamic Systems Theory) to explore pronunciation learning as a person-specific, cognitive, yet highly contextualized, social, and experiential phenomenon.

Other theoretical proposals offering promising avenues for conceptualizing L2 pronunciation development come from the field of social psychology. For example, the Willingness to Communicate framework developed by McIntyre and his colleagues (MacIntyre et al. 1998) incorporates a variety of cognitive, social, and experiential factors to explain a learner's choice to engage in communication in an L2. Similarly, Clément's Social Context model (Clément 1980; Clément and Kruidenier 1985) draws on such variables as L2 confidence, competence, and identity to describe intergroup contact. More recently, Clément, Baker, and MacIntyre (2003) provided empirical data supporting a framework based on a combination of these two models. In the combined model, frequency and quality of L2 contact predict L2 confidence, which is related to both willingness to communicate and identity. These two factors, in turn, both predict frequency of L2 use. It is certainly important to explore the applicability of these social-psychological models to L2 pronunciation development in different L2 learning contexts (e.g., classroom, naturalistic, study abroad). It is also important to investigate how such models might be used to explain various aspects of L2 pronunciation learning, for instance, variability in phonological development or acquisition of specific aspects of pronunciation (see Mougeon, Rehner, and Nadasdi 2004 for some work in this area).

Several multidimensional frameworks focus on L2 pronunciation learning specifically. For example, Segalowitz and his colleagues (Segalowitz, Gatbonton, and Trofimovich 2009) proposed a conceptualization of L2 pronunciation learning that includes several cognitive and social influences. In this framework, ethnic identity is part of a larger motivation system that determines whether and to what extent learners engage in L2 use. Language use is important because it provides learners with opportunities to tune their perceptual and cognitive systems for the processing of L2 input. This cognitive and perceptual tuning is driven by several psycholinguistic variables, which include frequency (i.e., how often a particular pronunciation target occurs in an L2) and cross-language similarity (i.e., perceptual differences between L1 and L2 that determine the ease or difficulty of certain aspects of L2 pronunciation). Thus, in this framework, ethnic identity and motivational variables shape particular patterns of L2 use. L2 use, in turn, impacts language learning outcomes by allowing learners to practise their cognitive processing skills through L2 input and/or output.

Moyer's (2004, 2009) integrated view of critical influences in L2 learning exemplifies another multidimensional framework relevant to understanding L2 pronunciation development. Moyer places learners' experience with L2 input, which she calls "strategic use of input", at the center of her framework. Strategic use of input refers to learners' choices in how and when they take advantage of the available input in accordance with their intentions, orientations, and cognitive styles. Moyer's framework also specifies several clusters of influences that shape how learners use language input. These clusters include cognitive influences (which involve instructional variables, learner strategies, and would also include attention to form), social influences (which encompass different language contact domains and situations of language use), and psychological influences (which involve attitudes, motivations, and identity issues). Moyer is deliberately vague in describing the precise contributions of these different factors to L2 pronunciation learning because these contributions are arguably specific to each learning context. At least one avenue for future research here would be to provide more refined descriptions of how different factors shape L2 pronunciation learning in specific learning contexts. This will allow researchers to use theoretical frameworks (such as the ones described by Segalowitz et al. and Moyer) not solely as descriptive tools but also as sources of empirically testable hypotheses.

Conclusion

We conclude our chapter with a quote from Eleanor Gibson, an American psychologist who, along with her husband James Gibson, developed a theory of human learning and development based on a complex interaction of people's cognitive abilities and environmental affordances, which refer to the possibilities for actions that a given environment offers. This view of learning assumes that development in early infancy proceeds as children experience various environments

(i.e., real-world contexts) and that children use perception to discover various affordances of such environments (e.g., reaching out for a moving object will afford the child to catch it). Through such experience, children both become more accurate at perceptual tasks and learn about the environment they are in. Applied to language, this conceptualization of perceptual learning is remarkably similar to the language-as-action view proposed by Clark (1992), where language use can be seen as perceiving and using affordances for speakers to accomplish real-world goals, making a change in a given social environment (e.g., asking a neighbor to turn down music). In one of her writings on affordances, Gibson noted that "the complementarity of the [human] and its environment is a whole and must be studied as such" and that "the more we try to decompose this complementarity by looking for elements, the more likely we are to sacrifice the meanings we are looking for" (Gibson 1991: 569). This quote aptly highlights the importance for researchers to study L2 pronunciation development as a complex, holistic phenomenon. It also underlines the danger of potentially missing important aspects by investigating L2 pronunciation learning as a function of individual, isolated variables. One important goal for future researchers is therefore to develop a multidimensional picture of L2 pronunciation learning as a complex sociocognitive and situationally embedded phenomenon.

REFERENCES

Abrahamsson, N. and Hyltenstam, K. 2008. The robustness of aptitude effects in near-native second language acquisition. *Studies in Second Language Acquisition* 30: 481–509.

Abrahamsson, N. and Hyltenstam, K. 2009. Age of acquisition and nativelikeness in a second language: listener perception versus linguistic scrutiny. *Language Learning* 59: 249–306. doi: 10.1111/j.1467-9922.2009.00507.x.

Aoyama, K., Flege, J.E., Guion, S.G., Akahane-Yamada, R., and Tsuneo Yamada, T. 2004. Perceived phonetic dissimilarity and L2 speech learning: the case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics* 32: 233–250. doi: 10.1016/S0095-4470(03)00036-6.

Aoyama, K., Guion, S.G., Flege, J.E., Yamada, T., and Akahane-Yamada, R. 2008. The first years in an L2-speaking environment: a comparison of Japanese children and adults learning American English. *International Review of Applied Linguistics* 46: 61–90.

Archila-Suerte, P., Zevin, J., Bunta, F., and Hernandez, A.E. 2012. Age of acquisition and proficiency in a second language independently influence the perception of non-native speech. *Bilingualism: Language and Cognition* 15: 190–201. doi: 10.1017/S1366728911000125.

Arellano, S.I. and Draper, J.E. 1972. Relations between musical aptitudes and second-language learning. *Hispania* 55: 111–121.

Ashmore, R.D., Deaux, K., and McLaughlin-Volpe, T. 2004. An organizing framework for collective identity: articulation and significance of multidimensionality. *Psychological Bulletin* 130: 80–114.

Atkinson, D. 2011. A sociocognitive approach to second language acquisition: how mind, body, and world work together

in learning additional languages. In: *Alternative Approaches to Second Language Acquisition*, D. Atkinson (ed.), 143–166, New York: Routledge.

Atkinson, D., Churchill, E., Nishino, T., and Okada, H. 2007. Alignment and interaction in a sociocognitive approach to second language Acquisition. *Modern Language Journal* 91: 169–188.

Au, T.K., Knightly, L.M., Jun, S.-A., and Oh, J.S. 2002. Overhearing a language during childhood. *Psychological Science* 13: 238–243. doi: 10.1111/1467-9280.00444.

Au, T.K., Oh, J.S., Knightly, L.M., Jun, S.-A., and Romo. L.F. 2008. Salvaging a childhood language. *Journal of Memory and Language* 58: 998–1011. doi: 10.1016/j.jml.2007.11.001.

Austin, J.L. 1975. Lecture I. In: *How to Do Things with Words (William James Lectures)*, J.O. Urmson and M. Sbisà (eds.), 1–11, Cambridge, MA: Harvard University Press.

Baker, W. and Trofimovich, P. 2005. Interaction of native- and second-language vowel system(s) in early and late bilinguals. *Language and Speech* 48: 1–27.

Baker, W., Trofimovich, P., Flege, J.E., Mack, M., and Halter, R. 2008. Child–adult differences in second-language phonological learning: the role of cross-language similarity. *Language and Speech* 51: 316–341. doi: 10.1177/0023830908099068.

Bentley, A.1966. *Measures of Musical Abilities*, London: Harrap.

Best, C.T. and Tyler, M.D. 2007. Nonnative and second-language speech perception: commonalities and complementarities. In: *Language Experience in Second Language Speech Learning*, O.-S. Bohn and M.J. Munro (eds.), 13–34. Amsterdam: John Benjamins.

Birdsong, D. 2009. Age and the end state of second language acquisition. In: *The New Handbook of Second Language Acquisition*, W.C. Ritchie and T.K. Bhatia (eds.), 401–424, Bingley, UK: Emerald.

Bohn, O.-S. and Flege, J.E. 1990. Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics* 11: 303–328. doi: 10.1017/S0142716400008912.

Bongaerts, T., van Summeren, C., Planken, B., and Schils, E. 1997. Age and ultimate attainment in the pronunciation of a foreign language. *Studies in Second Language Acquisition* 19: 447–465.

Bornstein, M.H. 1987. Sensitive period in development: definition, existence, utility, and meaning. In: *Sensitive Periods in Development*, M.H. Bornstein (ed.), 3–17, Hillsdale, NJ: Lawrence Erlbaum.

Carroll, J.B. and Sapon, S.1959. *The Modern Language Aptitude Test*, San Antonio, TX: Psychological Corporation.

Chomsky, N. 1957. *Syntactic Structures*, Gravenhage, Holland: Mouton.

Churchill, E., Nishino, T., Okada, H.,and Atkinson, D. 2010. Symbiotic gesture and the sociocognitive visibility of grammar. *Modern Language Journal* 94: 234–253.

Clark, H.H. 1992. *Arenas of Language Use*, Chicago, IL: University of Chicago Press.

Clément, R. 1980. Ethnicity, contact and communicative competence in a second language. In: *Language: Social Psychological Perspectives*, H. Giles, W.P. Robinson, and P.M. Smith (eds.), 147–154, Oxford: Pergamon.

Clément, R., Baker, S.C., and MacIntyre, P.D. 2003. Willingness to communicate in a second language: the effects of context, norms and vitality. *Journal of Language and Social Psychology* 22: 190–209.

Clément, R. and Kruidenier, B.G. 1985. Aptitude, attitude and motivation in second language proficiency: a test of Clément's model. *Journal of Language and Social Psychology* 4: 21–37.

Davidson, L. 2011. Phonetic and phonological factors in the second language production of phonemes and phonotactics. *Language and Linguistics Compass* 5: 126–139. doi: 10.1111/j.1749-818x.2010.00266.x.

de Bot, K., Lowie, W., and Verspoor, M. 2007. A dynamic systems theory approach to second language acquisition. *Bilingualism: Language and Cognition* 10: 7–21. doi: 10.1017/S1366728906002732.

Deci, E.L. and Ryan, R.M. 1985. *Intrinsic Motivation and Self-determination in Human Behavior*, New York: Plenum.

DeKeyser, R.M. 2012. Age effects in second language learning. In: *Handbook of Second Language Acquisition*, S.M. Gass and A. Mackey (eds.), 442–460, London: Routledge.

DeKeyser, R.M., Alfi-Shabtay, I., and Ravid, D. 2010. Cross-linguistic evidence for the nature of age effects in second language acquisition. *Applied Psycholinguistics* 31: 413–438. doi: 10.1017/S0142716410000056.

Derwing, T.M. 2003. What do ESL students say about their accents? *The Canadian Modern Language Review* 59: 547–566.

Derwing, T.M. and Munro, M.J. 2013. The development of L2 oral language skills in two L1 groups: a 7-year study. *Language Learning* 63: 163–185.

Derwing, T.M., Munro, M.J., and Thomson, R.I. 2008. A longitudinal study of ESL learners' fluency and comprehensibility development. *Applied Linguistics* 29: 359–380.

Dexter, E.S., and Omwake, K.T. (1934). The relation between pitch discrimination and accent in modern languages. *Journal of Applied Psychology* 18: 267–271.

Díaz-Campos, M. 2004. Context of learning in the acquisition of Spanish second language phonology. *Studies in Second Language Acquisition* 26: 249–273. doi: 10.1017/S0272263104262052.

Dörnyei, Z. and Ushioda, E. (eds.). 2009. *Motivation, Language Identity and the L2 Self*, Bristol: Multilingual Matters.

Ellinger, B. 2000. The relationship between ethnolinguistic identity and English language for native Russian speakers and native Hebrew speakers in Israel. *Journal of Multilingual and Multicultural Development* 21: 292–307.

Flege, J.E. 2002. Interactions between the native and second-language phonetic systems. In *An Integrated View of Language Development: Papers in Honor of Henning Wode*, T. Piske, A. Rohde, P. Burmeister, and H. Wode (eds.), 217–244, Verlag Trier: Wissenschaftlicher.

Flege, J.E. and Fletcher, K.L. 1992. Talker and listener effects on the perception of degree of foreign accent. *Journal of the Acoustical Society of America* 91: 370–389. doi: 10.1121/1.402780.

Flege, J.E., Frieda, E.M., and Nozawa, T. 1997. Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics* 25: 169–186.

Flege, J.E., Munro, M.J., and MacKay, I.R.A. 1995. Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America* 97: 3125–3134.

Flege, J.E., Takagi, N., and Mann, V. 1995. Japanese adults can learn to produce English /r/ and /l/ accurately. *Language and Speech* 38: 25–55.

Flege, J.E., Yeni-Komshian, G.H., and Liu, S. 1999. Age constraints on second language acquisition. *Journal of Memory and Language* 41: 78–104.

Flege, J.E., Birdsong, D., Bialystok, E., Mack, M., Sung, H., and Tsukada, K. 2006. Degree of foreign accent in English sentences produced by Korean children and adults. *Journal of Phonetics* 34: 153–175. doi: 10.1016/j.wocn.2005.05.001.

Gatbonton, E. and Trofimovich, P. 2008. The ethnic group affiliation and L2 proficiency link: empirical evidence. *Language Awareness* 17: 229–248.

Gibson, E.J. 1991. *An Odyssey in Learning and Perception*, Cambridge, MA: MIT Press.

Giles, H., Bourhis, R.Y., and Taylor, D.M. 1977. Towards a theory of language in ethnic group relations. In: *Language, Ethnicity and Intergroup Relations*, H. Giles (ed.), 307–348, New York: Academic Press.

Giles, H. and Powesland, P.F. 1975. *Speech Style and Social Evaluation*, London and New York: Academic Press.

Gordon, E.E. 1995. *Manual: Musical Aptitude Profile*, Chicago: GIA Publications.

Grigorenko, E., Sternberg, R.J., and Ehrman, M.E. 2000. A theory based approach to the measurement of foreign language learning ability: the Canal-F Theory and Test. *Modern Language Journal* 84: 390–405. doi: 10.1111/0026-7902.00076.

Guion, S.G., Flege, J.E., and Loftin, J.D. 2000. The effect of L1 use on pronunciation in Quichua-Spanish bilinguals. *Journal of Phonetics* 28: 27–42. doi: 10.1006/jpho.2000.0104.

Guion, S.G., Flege, J.E., Akahane-Yamada, R., and Pruitt, J.C. 2000. An investigation of current models of second language speech perception: the case of Japanese adults' perception of English consonants. *Journal of the Acoustical Society of America* 107: 2711–2724. doi: 10.1121/1.428657

Guiora, A.Z., Brannon, R.C.L., and Dull, C.Y. 1972. Empathy and second language learning. *Language Learning* 22: 111–130.

Hakuta, K., Bialystok, E., and Wiley, E. 2003. Critical evidence: a test of the critical-period hypothesis for second language acquisition. *Psychological Science* 14: 31–38. doi: 10.1111/1467-9280.01415.

Hansen, D. 1995. A study of the effect of the acculturation model on second language acquisition. In: *Second Language Acquisition Theory and Pedagogy*, F.R. Eckman, D. Highland, P.W. Lee, J. Mileham, and R.R. Weber (eds.), 305–316, Mahwah, NJ: Lawrence Erlbaum.

Jacobs, B. 1988. Neurobiological differentiation of primary and secondary language acquisition. *Studies in Second Language Acquisition* 10: 303–337. doi: 10.1017/S0272263100007476.

Jia, G. and Aaronson, D. 2003. A longitudinal study of Chinese children and adolescents learning English in the United States. *Applied Psycholinguistics* 24: 131–161. doi: 10.1017/S0142716403000079.

Kurata, N. 2010. Opportunities for foreign language learning and use within a learner's informal social networks. *Mind, Culture, and Activity* 17: 382–396.

Lamarre, P. 2013. Catching "Montréal on the Move" and challenging the discourse of unilingualism in Québec. *Anthropologica* 55: 41–56.

Lenneberg, E.H. 1967. *Biological Foundations of Language*, New York: John Wiley & Sons, Inc.

Lybeck, K. 2002. Cultural identification and second language pronunciation of Americans in Norway. *Modern Language Journal* 86: 174–191.

MacIntyre, P.D., Clément, R., Dörnyei, Z., and Noels, K.A. 1998. Conceptualizing willingness to communicate in a L2: a situational model of L2 confidence and affiliation. *The Modern Language Journal* 82: 545–562.

Martinsen, R.A., Baker, W., Dewey, D.P., Bown, J., and Johnson, C. 2010. Exploring diverse settings for language acquisition and use: comparing study abroad, service learning abroad, and foreign language housing. *Applied Language Learning* 20: 40–69.

Marx, N. 2002. Never quite a "native speaker": accent and identity in the L2 – and the L1. *Canadian Modern Language Review* 59: 264–281.

McCandliss, B.D., Fiez, J.A., Protopapas, A., Conway, M., and McClelland, J.L. 2002. Success and failure in teaching the r-l contrast to Japanese adults: predictions of a Hebbian model of plasticity and stabilization in spoken language perception. *Cognitive, Affective, and Behavioral Neuroscience* 2: 89–108.

Milovanov, R., Huotilainen, M., Välimäki, V., Esquef, P.A.A., and Tervaniemi, M. 2008. Musical aptitude and second language pronunciation skills in school-aged children: neural and behavioral evidence. *Brain Research* 1194: 81–89. doi: 10.1016/j.brainres.2007.11.042.

Milovanov, R., Pietilä, P., Tervaniemi, M., and Esquef, P.A.A. 2010. Foreign language pronunciation skills and musical aptitude: a study of Finnish adults with higher education. *Learning and Individual Differences* 20: 56–60. doi: 10.1016/j.lindif.2009.11.003.

Mougeon, R., Rehner, K., and Nadasdi, T. 2004. The learning of spoken French variation by immersion students from Toronto, Canada. *Journal of Sociolinguistics* 8: 408–432.

Moyer, A. 1999. Ultimate attainment in L2 phonology: the critical factors of age, motivation, and instruction. *Studies in Second Language Acquisition* 21: 81–108. doi: 10.1111/j.1467-1770.1994.tb00635.x.

Moyer, A. 2004. *Age, Accent and Experience in Second Language Acquisition*, Clevedon: Multilingual Matters.

Moyer, A. 2009. Input as a critical means to an end: quantity and quality of experience in L2 phonological attainment. In: *Input Matters in SLA*, T. Piske and M. Young-Scholten (eds.), 159–174, Bristol: Multilingual Matters.

Moyer, A. 2011. An investigation of experience in L2 phonology. *Canadian Modern Language Review* 67: 191–216. doi: 10.3138/cmlr.67.2.191.

Munro, M.J., and Derwing, T.M. (2008). Segmental acquisition in adult ESL learners: a longitudinal study of vowel production. *Language Learning* 58: 479–502.

Oyama, S. 1976. A sensitive period for the acquisition of a nonnative phonological system. *Journal of Psycholinguistic Research* 5: 261–283.

Palmer, C. and Hutchins, S. 2006. What is musical prosody? In: *Psychology of Learning and Motivation*, B.H. Ross (ed.), vol. 46: 245–278, Amsterdam: Elsevier Press.

Paradis, M. 2009. *Declarative and Procedural Determinants of Second Languages*, Amsterdam: John Benjamins.

Penfield, W. and Roberts, L. 1959. *Speech and Brain-Mechanisms*, Princeton, NJ: Princeton University Press.

Pickering, M.J. and Garrod, S. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences* 27: 169–226.

Piller, I. 2002. Passing for a native speaker: identity and success in second language learning. *Journal of Sociolinguistics* 6: 179–206.

Pimsleur, P. 1966. *The Pimsleur Language Aptitude Battery*, New York: Harcourt, Brace, Jovanovich.

Pimsleur, P., Stockwell, R.P., and Comrey, A.L. 1962. Foreign language learning ability. *Journal of Educational Psychology* 53: 15–26. doi: 10.1037/h0044336.

Polat, N. 2011. Gender differences in motivation and L2 accent attainment: an investigation of young Kurdish learners of Turkish. *The Language Learning Journal* 39: 19–41.

Polat, N. and Mahalingappa, L.J. 2010. Gender differences in identity and acculturation patterns and L2 accent attainment. *Journal of Language, Identity, and Education* 9: 17–35. doi: 10.1080/15348450903476832.

Pulvermüller, F. and Schumann, J.H. 1994. Neurobiological mechanisms of language acquisition. *Language Learning* 44: 681–734.

Saito, K. (2013). Age effects on late bilingualism: the production development of /ɹ/ by high-proficiency Japanese learners of English. *Journal of Memory and Language* 69: 546–562.

Seashore, C.E. 1919. *The Psychology of Musical Talent*, New York: Silver, Burdett.

Segalowitz, N. and Freed, B. 2004. Context, Contact, and Cognition in Oral Fluency Acquisition. *Studies in Second Language Acquisition* 26: 173–199. doi: 10.1017/S0272263104262027.

Segalowitz, N., Gatbonton, E., and Trofimovich, P. 2009. Links between ethnolinguistic affiliation, self-related motivation and second language fluency: Are they mediated by psycholinguistic variables?" In: *Motivation, Language Identity and the L2 Self*, Z. Dörnyei and E. Ushioda (eds.), 172–192, Bristol: Multilingual Matters.

Skehan, P. 2012. Language aptitude. In: *Handbook of Second Language Acquisition*, S.M. Gass and A. Mackey (eds.), 381–395, London: Routledge.

Slevc, L.R. and Miyake, A. 2006. Individual differences in second-language proficiency: Does musical ability matter? *Psychological Science* 17: 675–681. doi: 10.1111/j.1467-9280.2006.01765.x.

Sparks, R.L., Humbach, N., Patton, J., and Ganschow, L. 2011. Subcomponents of second-language aptitude and second-language proficiency. *Modern Language Journal* 95: 253–273. doi: 10.1111/j.1540-4781.2011.01176.x.

Strange, W. 2007. Cross-language phonetic similarity of vowels: theoretical and methodological issues. In: *Language Experience in Second Language Speech Learning*, O.-S. Bohn and M.J. Munro (eds.), 35–55, Amsterdam: John Benjamins.

Strange, W., Hisagi, M., Akahane-Yamada, R., and Kubo, R. 2011. Cross-language perceptual similarity predicts categorial discrimination of American vowels by naive Japanese listeners. *The Journal of the Acoustical Society of America* 130: EL226–EL231.

Tahta, S., Wood, M., and Loewenthal, K.1981. Foreign accents: factors related to transfer of accent from the first language to a second language. *Language and Speech* 24: 265–272. doi: 10.1177/002383098102400306.

Taylor, D.M., Meynard, R., and Rheault, E. 1977. Threat to ethnic identity and second language learning. In: *Language, Ethnicity and Intergroup Relations*, H. Giles (ed.), 99–116, New York: Academic Press.

Thompson, I. 1991. Foreign accents revisited: the English pronunciation of Russian immigrants. *Language Learning* 41: 177–204. doi: 10.1111/j.1467-1770.1991.tb00683.x.

Thomson, R.I. 2012. Improving L2 listeners' perception of English vowels: a computer-mediated approach. *Language Learning* 62: 1231–1258.

Trofimovich, P. (2013). Interactive alignment: a teaching-friendly view of second language pronunciation learning. *Language Teaching*. doi: 10.1017/S0261444813000360.

Trofimovich, P. and Baker, W. 2007. Learning prosody and fluency characteristics of L2 speech: the effect of experience on child learners' acquisition of five suprasegmentals. *Applied Psycholinguistics* 28: 251–276.

Trofimovich, P., McDonough, K., and Foote, J.A. (2014). Interactive alignment of multisyllabic stress patterns in a second language classroom. *TESOL Quarterly* 48: 815–832.

Trofimovich, P., McDonough, K., and Neumann, H. 2013. Using collaborative tasks to elicit auditory and structural priming. *TESOL Quarterly* 47: 177–186.

Trofimovich, P., Turuševa, L., and Gatbonton, E. 2013. Group membership and identity issues in second language learning. *Language Teaching* 46: 553–567.

Trofimovich, P., Lightbown. P.M., Halter, R., and Song, H. 2009. Comprehension-based practice: the development of L2 pronunciation in a listening and reading program. *Studies in Second Language Acquisition* 31: 609–639.

Ullman, M.T. 2005. A cognitive neuroscience perspective on second language acquisition: the declarative/procedural model. In: *Mind and Context in Adult Second Language Acquisition*, C. Sanz (ed.), 141–178, Washington, DC: Georgetown University Press.

Ushioda, E. and Dörnyei, Z. 2012. Motivation. In: *Handbook of Second Language Acquisition*, S.M. Gass and A. Mackey (eds.), 396–409, London: Routledge.

van Geert, P., Steenbeek, H., and van Dijk, M. 2011. A dynamic model of expert–novice co-adaptation during language learning and acquisition. In: *Modeling Bilingualism: From Structure to Chaos*, M.S. Schmid and W. Lowie (eds.), 235–266, Amsterdam: John Benjamins.

Wing, H.D. 1968. *Tests of Musical Ability and Appreciation: An Investigation into the Measurement, Distribution, and Development of Musical Capacity*, 2nd edition, London: Cambridge University Press.

Yager, K. 1998. Learning Spanish in Mexico: the effect of informal contact and student attitudes on language gain. *Hispania* 81: 898–913.

Yeni-Komshian, G., Flege, J.E., and Liu, S. 2000. Pronunciation proficiency in the first and second languages of Korean-English bilinguals. *Bilingualism: Language and Cognition* 3: 131–150.

Part VI Pronunciation Teaching

21 Intelligibility in Research and Practice: Teaching Priorities

MURRAY J. MUNRO AND
TRACEY M. DERWING

Introduction

Intelligibility has long been understood to be a fundamental requirement for effective human communication. Recognition of the centrality of this concept has resulted in a vast literature from such diverse fields as communication technology, audio engineering, speech language pathology, and audiology. In English language teaching, agreement about the importance of intelligibility is increasing among researchers and practitioners; however, as Levis (2005) points out, pronunciation teaching continues to be dominated by two contradictory principles: the nativeness principle and the intelligibility principle. The first of these assumes that learners should strive to become native-like in all aspects of pronunciation, whereas the intelligibility principle holds that learners should aim to develop speaking patterns that allow them to communicate with ease, even if their accent retains nonnative characteristics. The strong evidence suggesting that adult language learners rarely, if ever, achieve fully native-like pronunciation (Abrahamsson and Hyltenstam 2009; Flege, Munro, and MacKay 1995) and the compelling evidence that such a goal is unnecessary for effective communication (Munro and Derwing 1995a) have led us to work within a framework following from the intelligibility principle. Our attention below focuses on research about and applications of the intelligibility construct. In view of the fact that a fully satisfactory definition of the term intelligibility has proved elusive and no consensus exists on how to best measure it, we review the historical developments underpinning the current status of this notion.

The Handbook of English Pronunciation, First Edition. Edited by Marnie Reed and John M. Levis.

Definitions

As early as 1900, the famous phonetician, Henry Sweet, raised the issue of intelligibility as a central goal in the learning and teaching of languages. He was one of the first in a long line of British academics who emphasized intelligibility, including Abercrombie (1949), who argued that second language (L2) speakers do not need to emulate native speakers, and Gimson (1962), whose *Introduction to the Pronunciation of English* is still in print. This work encourages language learners to strive for intelligibility rather than native speaker-like productions. While these pronunciation specialists offered no technical definition of the concept, they evidently assumed a notion of intelligibility as an umbrella term meaning a shared understanding between speaker and listener. Voegelin and Harris (1951: 327), working from an anthropological perspective, proposed an even broader, cross-dialectal definition: the degree to which "people of one community understand the speech of another".

Attempts to formally define the intelligibility of L2 speech gradually emerged in the mid to late 20th century. However, the result was a diversity of opinions about how to describe the concept and how to evaluate it empirically. Catford (1950) distinguished the intelligibility of an utterance from its *effectiveness*, arguing that speech is intelligible when the listener identifies words correctly. An utterance is *effective* if the listener responds according to the speaker's intention. He did not, however, support this distinction with any phonological examples. Moreover, his conception of *unintelligible* appears to be tautological in that he described unintelligible speech as including speech that is intelligible but ineffective.

In their landmark article, Smith and Nelson (1985: 334) proposed a tripartite distinction among terms relating to understanding: "(1) intelligibility: word/utterance recognition; (2) comprehensibility: word/utterance meaning (locutionary force); (3) interpretability: meaning behind word/utterance (illocutionary force)." In this hierarchical framework, the third level was the highest. Although it seems essential to distinguish word recognition from other aspects of understanding, this model does not appear to have motivated much work on L2 communication, perhaps because it is unclear how to apply it empirically. For instance, a problem of defining intelligibility as the lowest level in a hierarchy of understanding is that high-level listener processing is sometimes invoked in identifying a word. That is, the comprehension of speech is not a linear process: when a word cannot be recognized on the basis of bottom-up cognitive processes, the listener may nonetheless be able to "fill it in" by exploiting top-down strategies. Thus a transcription task cannot be assumed to exclusively measure Smith and Nelson's (1985) notion of intelligibility because the task reflects more than just low-level speech processing.

A different approach was taken by Varonis and Gass (1982), who used the term "understanding" in the sense of "intelligibility" and described "comprehensibility" as a perceptual rating scale ranging from "I understood this sentence easily" to "I didn't understand it at all". In 1984, however, Gass and Varonis

operationalized "comprehensibility" differently, as the percentage of words correctly transcribed by listeners as they heard L2 speakers' productions; in other words, they used the word "comprehensibility" to refer to what many others have called "intelligibility".

A perspective on intelligibility from the speech sciences is that of Schiavetti (1992: 13): "the match between the intention of the speaker and the response of the listener to the speech passed through the transmission system". With respect to second language speech, Munro and Derwing (1995a: 76) characterized intelligibility as "the extent to which a speaker's message is understood by a listener". However, these definitions raise the issue of what we mean by such notions as "intention", "message", and "understand". While it seems obvious, for instance, that "understanding a spoken message" entails recognizing and grasping the meaning of most or all the individual words that the speaker has produced, it is unclear how much attention should be focused on paralanguage and the more subtle aspects of speakers' intentions, such as the unstated implications of their utterances, i.e., the illocutionary force that Smith and Nelson (1985) have called "interpretability". Our concept of intelligibility is broader and nonhierarchical, encompassing all three of Smith and Nelson's concepts, and recognizing that nonlinguistic factors, such as degree of shared knowledge and social context, may also affect understanding. Furthermore, it is not tied to any particular measurement—transcription tasks, comprehension questions, and other approaches are all ways of tapping into intelligibility, though no single task is fully satisfactory. While our approach conflates different aspects of understanding, it offers the advantage of relatively straightforward empirical assessment.

Rather than dwell on articulating a formal definition, we believe it is more fruitful to discuss intelligibility in terms of a number of functional properties that have been established in empirical research. The following apply only to spoken language communication:

- Intelligibility arises out of human interaction, in particular, the experience of one or more listeners as they process spoken material from an interlocutor. It therefore does not reside exclusively in either the speaker or listener.
- Intelligibility is a continuous phenomenon, such that the listener may understand all or none of the spoken material, as well as any intermediate amount. Furthermore, listeners may sometimes misjudge how much they have actually understood. They may realize, for instance, that some of the speaker's words are not intelligible, but they may also assume that they have understood other words, when, in fact, they have not. This potential for misapprehension complicates the assessment of intelligibility (see below).
- Intelligibility is affected by the speech transmission system (a telephone, the Internet, the air, water) as well as the ambient environment (quiet, noisy).
- Intelligibility is at least partially independent of many other commonly discussed dimensions of speech, such as accentedness, comprehensibility, fluency, accuracy, or naturalness.

Within our framework for L2 speech assessment, two concepts related to intelligibility are *comprehensibility* and *accentedness*. Munro and Derwing (1995a) and Derwing and Munro (1997) use these terms to mean ease or effort of understanding and degree of difference from some comparison pattern respectively. Each dimension has its own continuum, ranging from high to low. Although intelligibility and comprehensibility are more important than accentedness to language learners' communication skills, there is no hierarchical relationship among these dimensions. Comprehensibility and accentedness are partially independent of intelligibility, such that it is possible to be fully intelligible but somewhat difficult to understand. Furthermore, a speaker can be perceived to have a heavy accent, and yet be easy to understand and fully intelligible. The possible high–low combinations of intelligibility with comprehensibility are shown in Table 21.1, while the same combinations with accent appear in Table 21.2.

Table 21.1 Results of possible intelligibility and comprehensibility combinations.

Intelligibility	*Comprehensibility*	*Result*
High	High	Utterance is fully understood; little effort required
High	Low	Utterance is fully understood; great effort is required
Low	Low	Utterance is not (fully) understood; great effort is exerted
Low	High	Probably rare. Utterance is not fully understood; however, the listener has the false impression of having easily determined the speaker's intended meaning

Table 21.2 Results of possible intelligibility and accentedness combinations.

Intelligibility	*Accentedness*	*Result*
High	High	Utterance is fully understood; accent is very strong
High	Low	Utterance is fully understood; accent is barely noticeable
Low	Low	Not relevant to pronunciation; however, an utterance could be semantically anomalous, grammatically impossible, or obscured by noise and therefore unintelligible
Low	High	Utterance is not (fully) understood; accent is very strong

Local versus global intelligibility

An important distinction that is of use to both researchers and teachers is the difference between local and global intelligibility. The former refers to how well listeners recognize relatively small units of speech, such as segments and words, outside of a larger meaningful context, whereas the latter entails larger units of language that include rich contextual information. Although other researchers have not often used this terminology, previous empirical studies of intelligibility generally focus on one or the other. Field (2005), for instance, examined local intelligibility when he evaluated the effect of stress placement on listeners' identifications of isolated single words. As he points out, this type of identification task has a different locus than that of a sentence or narrative dictation in which the listener may exploit contextual information to assist understanding. The latter task is global and more closely approximates language interaction in real communicative situations. Ou, Yeh, and Chuang (2012) found large differences in intelligibility scores, depending on whether they used a local approach in which individual words were transcribed (43% incorrect) or a global approach in which sentences containing the same words of interest were included (12% incorrect). Within our framework, global intelligibility is the goal of pronunciation instructors who want to enhance their students' speech; research on local intelligibility is more useful to our understanding of L2 learning processes and to identifying some of the underlying components of global intelligibility. For instance, a local study might help us determine several speaker errors that lead to problems for the listener; however, only some of those may cause difficulties when contextual information is present. A fundamental problem for intelligibility researchers is to identify instances of the latter type. This concern is the one addressed by the functional load concept (Catford 1987). Phonological distinctions with a high functional load are those that "do a great deal of work" in the language. For instance, many minimal pairs in English are distinguished by the /i/ versus /ɪ/ contrast. Moreover, the words in such pairs are of high frequency and often belong to the same lexical classes. Consequently, one can predict, on a purely theoretical basis, that a failure to produce the phonemic distinction between the sounds is likely to result in significant miscommunication (see Brown 1991; Levis and Cortes 2008), perhaps even when detailed contextual information is available.

Measurement

The need for satisfactory assessment of intelligibility has long been recognized in the speech sciences. In reference to speech production by the deaf, for example, Subtelny (1977: 183) described intelligibility as "the most practical single index to apply in assessing competence in oral communication". Although a number of studies of L2 speech intelligibility have appeared in leading journals in recent years, relatively little attention has been directed to establishing the validity and reliability of intelligibility measures (Harding 2011). The fact that intelligibility is an aspect of

interaction, for instance, means that it can be assessed only by reference to listeners' experience. It is therefore not possible to measure it through an acoustic analysis of speech or through an expert phonetician's fine-grained impressionistic analysis. Nor can a listener be expected to directly assess the intelligibility of a speaker. For instance, it is not logically possible to collect ratings of intelligibility on a scale; in the absence of any corroborating evidence, listeners may accurately identify their own failure to understand, but may also mistakenly assume successful communication where a breakdown actually occurred. Intelligibility assessment is only possible if the speaker's intended utterance is known to the researcher and compared with the interpretation that the listener attributes to that same utterance. Given these constraints, intelligibility can be quantified in several ways, each of which has its own advantages and limitations. Here we will consider the following approaches that have been utilized in empirical work to assess intelligibility: word count, sentence verification, cloze and dictation, content summaries, and comprehension questions. We stress that each of these measures provides a window on the same, or closely related, underlying processes experienced by the listener. We therefore see them all as intelligibility measures, each of which is imperfect in its own way. A word count approach, for example, focuses strictly on exact word matches, but does not fully address illocutionary force, which would require further probing, perhaps with comprehension questions. On the other hand, using only comprehension questions entails a high risk of missing some aspects of a mismatch between the speaker's specific production and the listener's understanding because of the impossibility of evaluating every possible aspect of comprehension.

Open dictation with word count

By far the most common technique for measuring intelligibility is to have listeners transcribe utterances produced by an L2 speaker, and then count the number of correctly transcribed words (based on the speaker's intent). This requires that the researcher be certain of exactly all the intended words. This is guaranteed for controlled speaking tasks, such as a sentence or paragraph reading. For less constrained speech samples, it may be necessary for the researcher to question the speaker, or to listen to the utterances repeatedly until they become clear. A drawback of the word counting approach is the lack of a straightforward correspondence between the number of words correctly heard and the actual apprehension of the intended meaning of the full utterance. For instance, missing a critical word in a particular utterance may jeopardize the interpretation of the entire sentence. In some other utterance, however, misunderstanding several less critical words may have little or no negative effect on interpretation if the missing words can be inferred. A particular strength of dictation tasks, however, is a high degree of inter-listener reliability. Commonly, some utterances are consistently intelligible across listeners, while others are consistently unintelligible or partially unintelligible. Thus the data from word count tasks, despite some infelicities at a microlevel, are very useful when several listeners, speakers, and speech samples are employed (Munro and Derwing 1995a).

Cloze

A listening cloze is a less-demanding subset of a dictation task, for both listeners and researchers, because only certain words are the focus of the analysis. In this approach, listeners are presented with a written version of a spoken passage with certain vocabulary items deleted. Participants fill in the blanks while listening to the L2 speaker (e.g., Rubin 1992). This technique allows the researcher to establish the listener's comprehension of targeted discrete lexical items (which may have been chosen deliberately, based on the speaker's productions). A potential disadvantage is that the written text may provide contextual support to the listener, thus making the material more intelligible than it would be in its aural format (dictation).

Focused interviews of listeners

In her fine-grained analyses, Zielinski (2008) examined phonological contributors to unintelligibility by having listeners transcribe L2 utterances and then interviewing them with respect to each error. She could thus establish which patterns of L2 errors relating to L1 listening expectations most affected intelligibility. Although this technique yields detailed and accurate findings, it has the drawback of being extremely labor-intensive and unsuitable for use with large samples of listeners.

Sentence verification

Another technique for assessing L2 intelligibility is the sentence verification task, in which listeners judge the truth value of a set of utterances that can be readily evaluated from world knowledge (e.g., The inside of an egg is blue; Many people drink coffee at breakfast) (see Munro and Derwing 1995b). The listeners' response alternatives are *true*, *false*, and *not sure* (to discourage guessing). This technique requires that the speakers read a prepared list of true/false sentences; thus the speakers are constrained to use language that is not the product of their own linguistic competence.

Summaries

Hahn (2004) measured intelligibility of accented speech by asking listeners to recall as much of a minilecture as possible. After listening, the participants wrote as many of the main ideas and details as they could. Counts of correct main ideas proved to be useful in distinguishing different speakers' intelligibility, though counts of detail were not. Perlmutter (1989) also employed a recall task, focusing on major points for the same purpose. While summaries are useful for assessing comprehension at a broad level, they usually cannot provide detailed information about the actual locations of specific intelligibility breakdowns in a speaker's output.

Comprehension questions

Comprehension questions were also employed by Hahn (2004) to measure listeners' understanding of L2 speech. In that study, however, the listeners' scores were not significantly different despite variations in intelligibility that were established by the other measures. Thus, the approach appeared to be insufficiently sensitive.

Laboratory and classroom-based studies of intelligibility

It is tempting to make a sharp distinction between laboratory and classroom-based research on pronunciation. In fact, however, this dichotomy is artificial and fails to take into account the many ethical and scientific issues that arise in evaluating the effects of instruction on intelligibility. The fact that a study has been carried out in a classroom does not necessarily make its results any more generalizable to typical classrooms than a study carried out in a laboratory. While some studies are classroom-based in the sense that the participants are registered in existing classes and the research is integrated into their regular instruction, the nature of the control and the experimental procedures are often identical to those used in laboratory settings. For example, in Derwing, Munro, and Wiebe (1998), three existing classes of ESL learners participated under three distinct conditions (suprasegmental instruction, segmental instruction, no specific pronunciation instruction). Although instruction took place in the students' normal learning environment and was delivered by their homeroom teachers, their productions were collected and assessed using controlled laboratory methods. Saito and Lyster's (2012) study combined several features of classroom and laboratory environments in that participants were recruited (as in a lab study) to form instructional cohorts over a period of two weeks. The students received differential pronunciation teaching and were recorded under laboratory conditions; the resultant data were assessed using acoustic measurements. The finding that /ɹ/ productions were acoustically more native-like when form-focused instruction was used together with corrective feedback led the authors to conclude that the condition was effective in improving intelligibility. However, no direct measures of intelligibility were included in the study. Thus, studies fall on a continuum that ranges from more classroom-like to more laboratory-like, but a key issue for pedagogical purposes is how intelligibility is measured, whether a particular intervention is effective, and what the study indicates about methods and foci.

Although some L2 production studies are concerned with matters of intelligibility, others tend to emphasize the acquisition and consequences of pronunciation accuracy compared to a native speaker norm. Most pronunciation research studies fall into three broad categories: *focus-oriented* studies that attempt to identify and assess the impact of characteristics of second language speech detrimental to intelligibility; *acquisitional* studies that trace the development of pronunciation

intelligibility in L2 learners (using either cross-sectional or longitudinal designs); and *intervention* studies that examine whether a given method or technique is effective in bringing about changes in pronunciation.

Focus-oriented studies

Some focus-oriented studies use manipulated speech to evaluate listeners' comprehension in a controlled way. Tajima, Port, and Dalby (1997), for instance, modified timing in L2 speech samples to more closely approximate a native speaker model; listeners were then required to pick the phrases intended by the speakers in a multiple-choice task containing distractors developed from earlier comprehension problems. The modifications included deletions of epenthetic schwa, lengthening, and shortening of some segments, where appropriate, and additions of silence. Such changes were intended as improvements to the rhythmic properties of the original non-native utterances. The listeners' performance on the modified versus unmodified speech indicated that the temporal improvements led to increased intelligibility. This study is a useful contribution to our understanding of L2 speech intelligibility because it points to the probable benefits of working on rhythm as part of a pronunciation curriculum.

Another focus-oriented study with implications for L2 intelligibility is that of Hahn (2004), who investigated the role of primary sentence stress using three types of utterances in minilectures: correct productions, sentences with misassigned stress, and monotone (stressless) sentences. Listeners recalled more of the main ideas in the lectures when they heard appropriate stress assignment. Similarly, Field (2005) examined the effect of correct and incorrect word stress on lexical recognition. Listeners identified correctly stressed words better than words with misassigned stress; however, they also performed better on incorrectly stressed words when the inappropriately stressed vowel was produced with full quality (e.g., lagoon [lə'gu:n] produced as ['lægun]). The findings of Hahn's (2004) and Field's (2005) studies provide valuable information to L2 teachers who must make classroom decisions regarding their own students' intelligibility. In particular, stress influences intelligibility and should be taught if students regularly have difficulties with it.

Munro and Derwing (2006) conducted a preliminary study testing Catford's (1987) and Brown's (1991) hypotheses that functional load (the number of minimal pairs separated by two segments) should be taken into consideration when choosing segmental issues for classroom attention. Munro and Derwing found that low functional load errors (e.g., /f/ versus /θ/) had less effect on comprehensibility than did high functional load errors (e.g., /l / versus /n/), a result that has implications for the focus of pronunciation lessons.

The focus-oriented studies described above, although helpful in pointing to characteristics of L2 speech that interfere with intelligibility, do not address the issue of the extent to which such phenomena as rhythm, stress, and segments can be effectively taught; nor do they shed any light on suitable techniques. Rather, their outcomes must be used together with findings from intervention studies to ensure that learners are offered appropriate and effective intelligibility instruction.

Acquisition studies

A second source of information to guide teachers' decisions regarding which aspects of pronunciation should be taught comes from acquisition (i.e., noninstructed) studies of L2 pronunciation development. Research of this type clarifies the trajectories of learning that can be expected over time and therefore help in identifying problem areas that are unlikely to resolve themselves without intervention. Given the time constraints faced by English language instructors, it is imperative to make efficient and effective curriculum choices (Derwing 2008). For example, if learners are known to readily acquire a particular consonant easily there may be little or nothing to be gained by spending class time on that segment. Trofimovich and Baker (2006), in a cross-sectional study of Korean speakers learning English at the three-month, three-year, and ten-year points, found that stress-timing improved with English language experience, while other aspects of their speech, such as pause frequency, did not. On the basis of cross-sectional accentedness data from Mandarin adults living in the United States, Flege (1988) proposed that L2 pronunciation does not change much after the first year of experience in a new language environment. Some evidence in favor of that finding was obtained by Munro and Derwing (2008), who focused on vowel intelligibility. However, though vowel learning was indeed most rapid during the first year, additional improvements in vowel intelligibility were observed in the same speakers six years later. Just as Trofimovich and Baker (2006) found better performance on some suprasegmentals but not on others at ten years versus earlier points in time, Munro and Derwing's (2008) longitudinal data indicated that some aspects of Mandarin and Slavic language productions of English improved over time while others did not. For instance, they observed a significant improvement in /eɪ/ with near-native intelligibility by the end of the first year. Their performance on /ɪ/, however, was markedly different: both groups improved, but intelligibility was far below 100% after a full year of exposure. A subsequent study of the same speakers indicated that even after seven years, /ɪ/ productions were commonly unintelligible (Munro, Derwing, and Saito 2013). This suggests that spending a great deal of time teaching /eɪ/ is unnecessary, at least for groups from these language backgrounds, whereas some instruction on /ɪ/ could potentially be helpful, especially since the learner errors entailed confusion of /ɪ/ and /ɛ/, a high functional load pair. Further work will have to be carried out to determine how much benefit could be gained from such a focus. Analyses of /p/ productions by the Slavic language speakers from the same study indicated that some failed to learn to aspirate /p/ in the word initial position, even after seven years of residence in an English-speaking environment. The Mandarin speakers, in contrast, had no difficulty producing aspirated word-initial /p/, likely because Mandarin has aspirated stops. It is important to note, however, that even though problems with this consonant are somewhat tied to the learner's L1, not all Slavic speakers had difficulty producing intelligible aspirated /p/s. These preliminary studies are indicative of a need for more comprehensive research into naturalistic English language development to help guide pedagogical decisions. The outcomes

of these acquisitional studies provide evidence against a one-size-fits-all approach to pronunciation instruction. Moreover, they suggest that reliance on manuals or handbooks that attempt to characterize typical pronunciation errors based on L1 background (e.g., Swan and Smith 2001; Nilsen and Nilsen 2010) is ill-advised as a primary guide to L2 pronunciation curriculum design. In order to effectively address intelligibility issues, each learner must be individually assessed.

Intervention studies

A large number of segmental training studies have demonstrated improvement in perception of English consonant and vowel distinctions by L2 speakers. In most cases, these studies, often seen as prototypical laboratory research, were not designed with pedagogical implications in mind. However, there is no reason to automatically discount them as irrelevant to practical language learning. For instance, Thomson (2012a) has used principles from basic perception studies such as speaker variability (Thomson, Nearey, and Derwing 2009) to develop pedagogical software for individual practice of English segments. There is no question that, with feedback training, English learners can improve in perception and that improved perception can lead to some improvement in production (Bent and Bradlow 2003; Hayes-Harb et al. 2008). However, the limits of feedback training remain unclear in terms of how well the paradigms can be extended to suprasegmental phenomena, whether the training transfers to multiple contexts, and what limits exist on the levels of achievement that can be reached.

One of the earliest intervention studies of the effect of instruction on intelligibility was Perlmutter (1989), who examined ESL learners' performance before and after six months of instruction. Although Perlmutter concluded that the learners had benefited from instruction, the lack of a control group makes the validity of that claim unclear. Because previous work has already demonstrated that uninstructed L2 learners show improvement in intelligibility during the first year of residence, it is possible that the improvement would have occurred without instruction. Another intervention study conducted by Derwing, Munro, and Wiebe (1997) showed improvements in intelligibility in learners who had been living in an English environment for an average of ten years. That study also lacked a control group; however, because of the learners' length of residence, there was no reason to expect spontaneous improvement over the twelve-week period of the study. The Derwing, Munro, and Wiebe (1998) study cited above included a control group and compared two distinct intervention types (suprasegmental versus segmental) and their relative impacts on comprehensibility. Although both instructed groups showed post-intervention improvement when reading tasks were compared, only the supra-segmental group performed better in extemporaneous speech. Couper (2006), in an intervention study that did not assess intelligibility directly, found changes in the learners' speech that were attributable to instruction on avoiding epenthetic vowels and reducing consonant clusters. Since inaccurate syllable structure can result in a loss of intelligibility (Zielinski 2008), it is conceivable that Couper's participants became more intelligible as a result of

instruction. Derwing et al. (2014) also conducted a pronunciation intervention study with L2 speakers who had been in an English environment for an average of over nineteen years. Both intelligibility and comprehensibility improved, although in one task there was no change to accent. Furthermore, the L2 participants made perceptual gains over the course of the 12-week study. Because the pronunciation course covered several different aspects of speech, no direct connections between particular instructional techniques and improvement in intelligibility could be identified.

Listener effects

It may seem trite to repeat the well-known fact that oral communication is at minimum a two-person enterprise, in which speaker and listener have equal responsibility for ensuring a successful outcome. All intelligibility research therefore necessarily entails both speakers and listeners. Although the focus in L2 research is often on the characteristics of L2 speakers' productions, listeners play a crucial role in establishing the consequences of those characteristics. A number of researchers have pointed to the likely influence of listener variables on their comprehension of L2 speech (Munro 2008). It would not make sense to expect that every listener would respond in the same way to the same utterance. Rather, comprehension may vary depending upon a listener's familiarity with a given speaker, a specific accent, non-native accents in general, or with a particular topic (Gass and Varonis 1984; Kennedy and Trofimovich 2008). In some instances, sharing a common L1 is thought to enhance comprehension of other speakers; however, such effects appear to be very small and inconsistent (Munro, Derwing, and Holtby 2012; Munro, Derwing, and Morton 2006). Differences in listener aptitude can also be related to intelligibility, such that some listeners appear to have an affinity for L2 accented speech (Munro, Derwing, and Holtby 2012). Moreover, some listeners may feel more motivated and able to interact with an L2 speaker than others, and thus make greater efforts to understand (Derwing, Munro, and Thomson 2008; Derwing, Rossiter, and Munro 2002). The listener's age also influences understanding of L2 accented speech: elderly listeners appear to be at a disadvantage, even when their hearing is within normal limits for their age (Burda et al. 2003). Children also seem to understand less of L2 accented speech than adults (Munro, Derwing, and Holtby 2012). With all this potential variability, one has to ask whether teaching pronunciation is a viable enterprise. If listeners differ dramatically from each other in the ways in which they react to accented speech, how can an instructor decide on a focus for teaching? Correcting a particular problem might benefit some listeners but have no effect on others. Fortunately, despite the variability across listeners, evidence from rating studies reveals that their listening behavior is more similar than it is different. Presumably listeners process speech in very similar ways. If this were not the case, researchers would not see much interlistener agreement on which speech samples are intelligible and which are not. However, many studies show that diverse groups of

listeners agree on which L2 speakers from a set are easy or difficult to understand (comprehensibility) and the same is true of listeners' actual comprehension (intelligibility).

Teaching priorities in intelligibility-oriented instruction

In this section, we discuss priorities for pronunciation instruction with the understanding that language teachers require at least a basic knowledge of introductory linguistics and of the principles of L2 pronunciation in order to provide effective pedagogy. However, survey research indicates that many English language teachers have not actually received formal training in these areas (Foote, Holtby, and Derwing 2011).

Our conception of "prioritized pronunciation instruction" emphasizes helping learners to produce output that is comfortably intelligible (Abercrombie 1949), but is not concerned with "reducing" foreign accents. A first step in setting priorities within such a framework is to consider the kinds of changes in L2 speech production that might be effected through instruction. Table 21.3 lists eight logically possible outcomes of teaching interventions with respect to intelligibility, comprehensibility, and accentedness. A check mark in a cell indicates improvement on the relevant dimension, whereas an X indicates no change. Here we do not address the issue of whether some aspects of pronunciation might actually become worse as a function of instruction, though we recognize that such a risk exists.

Our interpretation of each combination is based on its consequences for a listener's understanding of the speaker's intended message. First, we consider simultaneous improvement of intelligibility and comprehensibility to be optimal, irrespective of whether the learner is heard as less accented. Second, any outcome in which one of intelligibility or comprehensibility improves is positive. It should be noted that the absence of improvement in one of these dimensions may at times indicate less than a fully desirable level of achievement; however, in other cases it may be an indication of a high level of performance on that dimension prior to instruction. For instance, a speaker who is already highly intelligible (and therefore cannot become more intelligible) could conceivably become more comprehensible thanks to instruction. In such a case, the speaker would have been highly intelligible *with effort* prior to instruction and highly intelligible *with less effort* afterwards. In short, the benefit for communication is that the listener must work less hard to understand the speaker. Third, when neither intelligibility nor comprehensibility improves, the outcome is negative, even if accent is reduced. This is because accent reduction is not relevant in prioritized pronunciation teaching, and should not be considered an appropriate goal when classroom time for instruction is limited. Of course, learners who wish to modify their pronunciation simply to approach some model are free to do so as they please.

Table 21.3 Possible outcomes of prioritized instruction.

Pattern	*Intelligibility*	*Comprehensibility*	*Accentedness*	*Interpretation*
1	✓	✓	✓	Optimal. Outcomes 1
2	✓	✓	X	and 2 are equivalent. Not only are the speakers' utterances understood more fully, but they are processed more easily by the listener.
3	✓	X	✓	Positive. The listener understands more of what the speaker has said, but experiences no greater ease in processing.
4	X	✓	✓	Positive. The listener experiences greater ease in processing the speaker's utterances, but does not understand more material.
5	✓	X	X	Positive. The listener understands more of what the speaker has said, but experiences no greater ease in processing; nor does the speech sound noticeably closer to native-like.
6	X	✓	X	Positive. The listener experiences greater ease in processing the speaker's utterances, but does not understand more material; nor does the speech sound noticeably closer to native-like.
7	X	X	✓	Negative. Outcomes 7
8	X	X	X	and 8 are equivalent. There is no change in the amount understood or ease of processing.

The possibilities in Table 21.3 are based on the assumption that the three dimensions in question are independent enough of each other that a change in one does not automatically entail a change in either of the others. It is important to note that such an assumption is not merely a theoretical conjecture. Rather, empirical support for the actual occurrence of most of them has already been obtained. In an intervention study, for instance, Derwing, Munro, and Wiebe (1997) observed individual instances of Patterns 1, 4, 5, 6, 7, and 8. Intriguingly, each of these patterns was observed in the same group of learners receiving the same type of instruction. Further work to account for why the individual learners responded differently to the same instruction is clearly needed. Additional studies of learners of English and of other languages support the independence of some of the two-way combinations as well. For example, Derwing, Munro, and Wiebe (1998) reported improved comprehensibility of L2 speakers' oral narratives with no change in accentedness, as well as improved accentedness in sentence productions without improved comprehensibility. Furthermore, when Holm (2008) provided intonation training to learners of Norwegian, she observed improvement in accent without improvement in intelligibility.

Focus priorities

Given the various possible instructional outcomes described in the last section, it is useful to consider the ways in which specific priorities can be identified to maximize optimal and positive outcomes. The first column of Table 21.4 provides a nonhierarchical list of six focus priorities for the promotion of global intelligibility and comprehensibility. Priority 1 is an emphasis on local phonological structures

Table 21.4 Teaching for global intelligibility and comprehensibility: priorities and implementation.

Focus priority	*Implementation*
(1) Emphasis on local phonological structures that enhance global intelligibility and comprehensibility	(1) Effective, efficient classroom management
(2) Priorities supported by empirical evidence	(2) Appropriate attention to discourse, utterance, and word levels
(3) Priorities based on sound theoretical grounds	(3) Satisfactory balance of communicative and formS-focused activities
(4) Emphasis on problems that do not resolve on their own	
(5) Coverage of errors shared by most students in class	
(6) Individualized assessment	

that enhance global intelligibility and comprehensibility. This means that not all local pronunciation problems deserve equal attention, especially since teachers usually have limited time to devote to pronunciation issues.

A problem we face at this point is the limited evidence about what those structures are. Our view is that teachers must base their curricula on the best available evidence about phonological phenomena that influence intelligibility and comprehensibility. In some cases, this may derive from well-constructed, controlled empirical studies (Priority 2), but in others it will come from Priority 3 – a well-motivated, theoretical framework. To ensure the most effective use of time, it is best to direct the most attention to pronunciation problems that are unlikely to resolve themselves in the long run without explicit intervention (focus Priority 4) and to devote class time to difficulties that are shared by many or all students in the class (focus Priority 5). On the latter point, suprasegmentals, which are often problematic for ESL learners from a range of L1 backgrounds, are good candidates for whole-class attention. Finally, it is essential that instructors evaluate individual students' speech to identify idiosyncratic patterns that interfere significantly with intelligibility. These may be addressed by individualized activities, either through technology or in small groups with the teacher.

Implementation

The second column of Table 21.4 lists means for the implementation of focus priorities, whether in a stand-alone pronunciation class or an integrated skills class. Teachers often find that exclusively lock-step, teacher-centered formats do not lend themselves well to many aspects of pronunciation teaching. For this reason, careful classroom management is important for successful instruction. While shared problems may be addressed through whole-class activities, for idiosyncratic difficulties, students may benefit more from rotating through several work stations in the classroom, completing work independently or with each other, and spending some of that time directly with the teacher in a small group of students who share similar problems. In addition, available technology makes it much easier for teachers to assign beneficial self-administered pronunciation tasks (e.g., Thomson's English Accent Coach 2012b; University of Iowa's *Phonetics: The Sounds of American English*, n/d). A second concern in implementation is the need for attention to units of language at multiple levels. In particular, it is essential that learners gain practice and receive feedback on discourse-level productions, as well as on shorter utterances and word-level language. Work on discourse is particularly important in the teaching of intonation, an aspect of English phonology that is essential to spoken-language effectiveness. Levis and Pickering (2004) point out the value of evolving speech visualization technologies such as pitch displays for instruction. In particular, this type of software allows learners to record and compare their productions visually with models to facilitate acquisition of such intonational phenomena as paratones and tonal composition.

On a third point, a common critique of contemporary pronunciation teaching is that it is excessively formS-focused as opposed to form-focused. This opposition of

approaches is not restricted to pronunciation, but has been widely discussed in the second language acquisition literature in connection with grammar (Ellis 2008). While formS-focused instruction draws students' attention to specific structural details of decontextualized language, form-focused instruction addresses language issues within meaningful interaction; typically students are encouraged to "discover" regularities within the language. Despite several investigations of grammar acquisition, relatively little research has addressed the use of either formS or forms in the area of L2 pronunciation. Moreover, we cannot assume that findings pertaining to grammar are automatically applicable to pronunciation. It is clear, however, that L2 speakers who do not receive any explicit pronunciation instruction in their ESL training frequently show minimal pronunciation development over time, regardless of degree of exposure. For this reason, we advocate a balanced approach that includes some focus on formS that is critical for intelligibility, particularly to guide learners in developing the articulatory skills needed to produce new segments and prosody. However, to ensure transfer to real-world communication, a focus on formS needs to be matched with sufficient practice during communicative activities in which learners must cope with the higher cognitive demands of producing original output.

Conclusions

In attempting to outline a detailed approach to prioritizing issues for the pronunciation classroom, we continually face the problem of limited data and the consequent need to speculate rather than provide empirically grounded bases for recommendations to teachers. Progress has been made in understanding how intelligibility relates to other constructs in L2 speech. In addition, the growing body of empirical research on pronunciation has shed some light on how the most serious pronunciation problems can be identified and effectively addressed. However, much research remains to be done. In particular, further testing of theoretical notions such as functional load is needed to develop a clearer picture of the relationship between local and global intelligibility. In addition, we currently still suffer from a dearth of intervention studies whose findings can be directly applied in classrooms. Far more work of this type is needed to help us understand the diverse ways in which individual learners respond to pronunciation instruction and to help us determine how to efficiently and effectively use class time to address shared and individual problems. Finally, the potential of technology for pronunciation instruction has yet to be effectively tapped. Regrettably, one of the biggest problems in the use of much currently available software is that it comes without priority-setting features. The common one-size-fits-all approach in which practice is offered in "everything" is unhelpful to teachers and students who need to focus their attention on issues that will genuinely improve their communication skills. An important challenge, then, is to find ways to apply the individualized attention that technology offers so that time is not wasted and interactional benefits are maximized.

REFERENCES

Abercrombie, D. 1949. Teaching pronunciation. *English Language Teaching* 3: 113–122.

Abrahamsson, N. and Hyltenstam, K. 2009. Age of onset and nativeness in a second language: listener perception versus linguistic scrutiny. *Language Learning* 59(2): 249–306.

Bent, T. and Bradlow, A. 2003. The interlanguage speech intelligibility benefit. *Journal of the Acoustical Society of America* 114: 1600–1610.

Brown, A. 1991. Functional load and the teaching of pronunciation. In: *Teaching English Pronunciation: A Book of Readings*, A. Brown (ed.), 211–224, London: Routledge.

Burda, A.N., Scherz, J.A., Hageman, C.F., and Edwards, H.T. 2003. Age and understanding of speakers with Spanish or Taiwanese accents. *Perceptual and Motor Skills* 97: 11–20.

Catford, J.C. 1950. Intelligibility. *English Language Teaching Journal* 1: 7–15.

Catford, J.C. 1987. Phonetics and the teaching of pronunciation: a systemic description of English phonology. In: *Current Perspectives on Pronunciation*, J. Morley (ed.), 878–100, Washington, DC: TESOL.

Couper, G. 2006. The short and long-term effects of pronunciation instruction. *Prospect* 21: 46–66.

Derwing, T.M. 2008. Curriculum issues in teaching pronunciation to second language learners. In: *Phonology and Second Language Acquisition*, J.H. Edwards and M. Zampini (eds.), 347–369. Amsterdam: John Benjamins.

Derwing, T.M. and Munro, M.J. 1997. Accent, intelligibility, and comprehensibility: evidence from four L1s. *Studies in Second Language Acquisition* 19: 1–16.

Derwing, T.M., Munro, M.J., and Thomson, R.I. 2008. A longitudinal study of ESL learners' fluency and comprehensibility development. *Applied Linguistics* 29(3): 359–380.

Derwing, T.M., Munro, M.J., and Wiebe, G.E. 1997. Pronunciation instruction for 'fossilized' learners: Can it help?" *Applied Language Learning* 8(2): 217–235.

Derwing, T.M., Munro, M.J., and Wiebe, G.E. 1998. Evidence in favor of a broad framework for pronunciation instruction. *Language Learning* 48(3): 393–410.

Derwing, T.M., Rossiter, M.J., and Munro, M.J. 2002. Teaching native speakers to listen to foreign-accented speech. *Journal of Multilingual and Multicultural Development* 23: 245–259.

Derwing, T.M., Munro, M.J., Foote, J.A., Waugh, E., and Fleming, J. 2014. Opening the window on comprehensible pronunciation after 19 years: a workplace training study. *Language Learning* 64: 526–548.

Ellis, R. 2008. *The Study of Second Language Acquisition*, 2nd Edition, Oxford: Oxford University Press.

Field, J. 2005. Intelligibility and the listener: the role of lexical stress. *TESOL Quarterly* 39: 399–423.

Flege, J.E. 1988. Factors affecting degree of perceived foreign accent in English sentences. *Journal of the Acoustical Society of America* 84: 70–79.

Flege, J.E., Munro, M.J., and MacKay, I. 1995. Factors affecting degree of perceived foreign accent in a second language. *Journal of the Acoustical Society of America* 97: 3125–3134.

Foote, J.A., Holtby, A., and Derwing, T.M. 2011. Survey of pronunciation teaching in adult ESL programs in Canada, 2010. *TESL Canada Journal* 29: 1–22.

Gass, S. and Varonis, E. 1984. The effect of familiarity on the comprehensibility of nonnative speech. *Language Learning* 34: 65–89.

Gimson, A.C. 1962. *Introduction to the Pronunciation of English*, London: Edward Arnold.

Hahn, L. 2004. Primary stress and intelligibility: research to motivate the teaching of suprasegmentals. *TESOL Quarterly* 38(2): 201–223.

Harding, L. 2011. *Accent and Listening Assessment: A Validation Study of the Use of peakers with L2 Accents on an Academic English Listening Test*. Frankfurt: Peter Lang.

Hayes-Harb, R., Smith, B.L., Bent, T., and Bradlow, A.R. 2008. The interlanguage speech intelligibility benefit for native speakers of Mandarin: production and perception of English word-final voicing contrasts. *Journal of Phonetics* 36: 664–679.

Holm, S. 2008. Intonational and durational contributions to the perception of foreign-accented Norwegian: An experimental phonetic investigation. Unpublished PhD dissertation, Department of Language and Communication Studies, Norwegian University of Science and Technology.

Kennedy, S. and Trofimovich, P. 2008. Intelligibility, comprehensibility, and accentedness of L2 speech: the role of listener experience. *The Canadian Modern Language Review* 64: 459–489.

Levis, J.M. 2005. Changing concepts and shifting paradigms in pronunciation teaching. *TESOL Quarterly* 39(3): 369–377.

Levis, J.M. and Cortes, V. 2008. Minimal pairs in spoken corpora: implications for pronunciation assessment and teaching. In: *Towards Adaptive CALL: Natural Language Processing for Diagnostic Language Assessment*, C.A. Chapelle, Y.-R. Chung, and J. Xu (eds.), 197–208, Ames, IA: Iowa State University.

Levis, J. and Pickering, P. 2004. Teaching intonation in discourse using speech visualization technology." *System* 32: 505–524.

Munro, M.J. 2008. Foreign accent and speech intelligibility. In: *Phonology and Second Language Acquisition*, J.G.H. Edwards and M.L. Zampini (eds.), 193–218, Amsterdam: John Benjamins.

Munro, M.J. and Derwing, T.M. 1995a. Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning* 45(1): 73–97.

Munro, M.J. and Derwing, T.M. 1995b. Processing time, accent, and comprehensibility in the perception of foreign-accented speech. *Language and Speech* 38(3): 289–306.

Munro, M.J. and Derwing, T.M. 2006. The functional load principle in ESL pronunciation instruction: an exploratory study. *System* 34: 520–531.

Munro, M.J. and Derwing, T.M. 2008. Segmental acquisition in adult ESL learners: a longitudinal study of vowel production. *Language Learning* 58: 479–502.

Munro, M.J., Derwing, T.M., and Holtby, A.K. 2012. Evaluating individual variability in foreign accent comprehension. In: *Proceedings of the 3rd Pronunciation in Second Language Learning and Teaching Conference,* September 2011, J. Levis and K. LeVelle (eds.), Ames, IA: Iowa State University.

Munro, M.J., Derwing, T.M., and Morton, S. 2006. The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition* 28: 111–131.

Munro, M.J., Derwing, T.M., and Saito, K. 2013. English L2 vowel acquisition over seven years. In: *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference*, J. Levis and K. LeVelle (eds.), August 2012, Ames, IA: Iowa State University.

Nilsen, D.L.F. and Nilsen, A.P. 2010. *Pronunciation Contrasts in English*, 2nd edition, Long Grove, IL: Waveland Press.

Ou, S.-C., Yeh, R., and Chuang, Z.-L. 2012. Units of analysis, intelligibility evaluation and phonological cores of EIL. Poster presented at *The 4th Annual Conference on Pronunciation in Second Language Learning and Teaching*, Vancouver.

Perlmutter, M. 1989. Intelligibility rating of L2 speech pre- and post-intervention. *Perceptual and Motor Skills* 68: 515–521.

Rubin, D.L. 1992. Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher Education* 33(4): 511–531.

Saito, K. and Lyster, R. 2012. Effects of form-focused instruction and corrective feedback on L2 pronunciation development of /ɹ/ by Japanese learners of English. *Language Learning* 62: 595–633. doi: 10.1111/j.1467-9922.2011.00639.x.

Schiavetti, N. 1992. Scaling procedures for the measurement of speech intelligibility. In: *Intelligibility in Speech Disorders*, R.D. Kent (ed.), 11–34, Philadelphia, PA: John Benjamins.

Smith, L.E. and Nelson, C.E. 1985. International intelligibility of English: directions and resources. *World Englishes* 3: 333–342.

Subtelny, J.D. 1977. Assessment of speech with implications for training. In: *Childhood Deafness: Causation, Assessment, and Management*, F.H. Bess (ed.), 183–194, New York: Grune & Stratton.

Swan, M. and Smith, B. (eds.). 2001. *Learner English: A Teacher's Guide to Interference and Other Problems*, 2nd edition, Cambridge: Cambridge University Press.

Sweet, H. 1900. *The Practical Study of Languages: A Guide for Teachers and Learners*, New York: Holt.

Tajima, K., Port, R., and Dalby, J. 1997. Effects of temporal correction on intelligibility of foreign accented English. *Journal of Phonetics* 25: 1–24.

Thomson, R.I. 2012a. Improving L2 listeners' perception of English vowels: a computer-mediated approach. *Language Learning*. doi: 10.1111/j.1467-9922.2012.00724.x.

Thomson, R.I. 2012b. *English Accent Coach*. Accessed October 24, 2012, http://www.englishaccentcoach.com.

Thomson, R.I., Nearey, T.M., and Derwing, T.M. 2009. A modified statistical pattern approach to measuring the crosslinguistic similarity of Mandarin and English vowels. *Journal of the Acoustical Society of America* 126: 1447–1460.

Trofimovich, P. and Baker, W. 2006. Learning second language suprasegmentals: effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition* 28: 1–30.

University of Iowa. n/d. *Phonetics: The Sounds of Spoken English*.http://www.uiowa.edu/~acadtech/phonetics/ Accessed July 9, 2013.

Varonis, E.M. and Gass, S. 1982. The comprehensibility of non-native speech. *Studies in Second Language Acquisition* 4: 114–136.

Voegelin, C.F. and Harris, Z.S. 1951. Methods for determining intelligibility among dialects of natural languages. *Proceedings of the American Philosophical Society* 95: 322–329.

Zielinski, B.W. 2008. The listener: no longer the silent partner in reduced intelligibility. *System* 36(1): 69–84.

22 The Segmental/ Suprasegmental Debate

BETH ZIELINSKI

Introduction

An important focus of ongoing research into L2 English pronunciation learning and teaching is the identification of features of pronunciation that impact on a speaker's intelligibility and comprehensibility.[1] Such features are generally identified and categorized by researchers as either segmental (individual sounds) or suprasegmental (extending over more than an individual sound, e.g., syllable structure, stress, rhythm, intonation). A long-standing debate in pronunciation teaching is whether segmentals or suprasegmentals are more important in promoting understandable speech. On one side of the debate, various authors have claimed that suprasegmental features should be given priority in pronunciation teaching because they are more important than segmental features to intelligibility and comprehensibility. For example, Fraser (2001: 33) listed six pronunciation features in the order in which they should be taught, based on their relative impact on listeners' comprehension. At the top of the list was *word and sentence stress* and the features further down the list involved consonant and vowel production and distinctions. Fraser stated that this order implies that stress is the most important thing to teach, as learners with perfect consonant distinctions will still be very difficult to understand if they have not mastered word and sentence stress. She argued that "there is little point in helping students with, say consonant distinctions, if they have very poor control of word and sentence stress" (2001: 34). Chela-Flores (2001) expressed a similar view when describing what she saw as the priority in her approach to teaching pronunciation: "More emphasis has been given to suprasegmental aspects, since these have more impact on intelligibility and help students with their immediate pronunciation needs" (2001: 98). Similarly, Tanner and Landon (2009) proposed that "if intelligibility is prioritized above accuracy, a focus on key words, stress, rhythm, and intonation rather than the articulation of individual sounds, may be needed" (2009: 51).

The Handbook of English Pronunciation, First Edition. Edited by Marnie Reed and John M. Levis.

On the other side of the debate, it has been argued that segmental features are more important to intelligibility and should therefore be given priority in pronunciation teaching. Collins and Mees (2003: 209), for example, supported this view and listed six pronunciation features they identified as having the greatest influence on intelligibility and therefore the highest priority in pronunciation teaching; five of these involved different consonants and vowels, and the sixth (and only suprasegmental feature) was word stress. Jenkins (2000) argued that segmental features are more important when non-native speakers of English are communicating with each other. She not only stressed the importance of segmental features to intelligibility in this context but asserted that some suprasegmental features actually "obstruct intelligibility" (2000: 135). She proposed a Lingua Franca Core, a set of pronunciation features considered to be crucial to intelligibility and thus a priority in pronunciation teaching. This set of pronunciation features was grouped into five categories referred to as "main core items" (see Jenkins 2002: 96–97 for a summary). Four of these categories involved segmental features, such as the production of various consonants, phonetic requirements related to voiced and voiceless consonants (aspiration and vowel length in specific contexts), production of consonant clusters, and production of specific vowels. Only one involved a suprasegmental feature, and this was the appropriate production and placement of nuclear stress, that is, stressing a particular word in an utterance to signal a particular meaning (variously referred to as tonic, primary, or contrastive stress). Other suprasegmental features, such as word stress, weak forms, stress-timed rhythm, and intonation, were considered to be non-core features, that is, not crucial to intelligibility.

Central to the segmental/suprasegmental debate is the notion that segmental and suprasegmental features are separate entities, and this is reflected in related research, where various studies have investigated the importance of one or the other to intelligibility and/or comprehensibility. Rogers and Dalby (2005), Bent, Bradlow, and Smith (2007), and Munro and Derwing (2006), for instance, focused on the relationship between intelligibility and/or comprehensibility and the production of various segments. Rogers and Dalby's findings highlight the importance of accurate vowel production to intelligibility, while Bent, Bradlow, and Smith found that both vowel accuracy and the accurate production of consonants in the word-initial position were important. Munro and Derwing used the theoretical concept of *functional load* to determine the impact of different consonant substitutions on listener judgments of comprehensibility, and found that those with a high functional load had a greater impact on comprehensibility judgments than those with a low functional load. The concept of functional load is based on the premise that some segmental contrasts do a greater amount of work in English than others, and are therefore more important to intelligibility and/or comprehensibility (see, for example, Brown 1991; Catford 1987; Gilner and Morales 2010).

Other studies have focused on the importance of suprasegmental features. Benrabah (1997), for example, found that non-target-like word stress was detrimental to intelligibility and Hahn (2004) found that both misplaced and no primary stress (i.e., nuclear stress) in a lecture impacted negatively on listener judgments of

the comprehensibility (ability to hear and understand) of the instructor. In contrast, Kang (2010) looked at the contribution of a range of different suprasegmental features (speech rate, pauses, stress, and pitch range) to listeners' judgments of comprehensibility but found that these judgments were more closely related to speech rate than to the other features.

Various other studies have investigated the importance of both segmental and suprasegmental features to intelligibility and/or comprehensibility and again have viewed and measured these features as separate entities. Munro and Derwing (1995) and Derwing and Munro (1997), for example, focused on the impact of both segmental and suprasegmental features on measures of intelligibility and comprehensibility. Munro and Derwing (1995) identified and counted non-target-like segments, but rated intonation on a 9-point scale where 1 = native-like and 9 = not at all native-like. Similarly, Derwing and Munro (1997) identified and counted non-target-like segments, but evaluated nativeness of prosody using a 9-point scale for prosodic goodness, where 1 = perfectly native-like and 9 = extremely accented. Furthermore, the samples rated for prosodic goodness had been filtered so that most of the segmental information had been removed, leaving rhythm and intonation intact. Derwing and Munro argued that in this way prosody could be assessed without the influence of segmental factors, that is, as separate from segmental features. In another study, Isaacs and Trofimovich (2012) investigated the influence of (amongst other linguistic features) both segmental and suprasegmental features on listener judgments of comprehensibility, and listed them as separate entities and measured them in different ways. The segmental features investigated were consonant and vowel production, and the suprasegmental features included syllable structure, word stress, vowel reduction related to rhythm, pitch contour, and pitch range.

The notion that segmental and suprasegmental features are separate entities is also central to the small number of empirical teaching studies investigating the impact on intelligibility and/or comprehensibility of teaching that focuses on different pronunciation features. The outcome of a teaching focus on segmental features (Saito 2011) and suprasegmental features (Tanner and Landon 2009) has been investigated, as well as the relative impact of segmental versus suprasegmental features (Derwing, Munro, and Wiebe 1997, 1998).

Despite the discussion and debate in the literature about the relative importance of segmental and suprasegmental features for intelligibility and comprehensibility, and which should be given priority in pronunciation teaching, there is little empirical evidence to support one over the other (see Derwing and Munro, 2005; Derwing and Munro, 2009; Levis, 2005). In fact, the general consensus seems to be moving towards the idea that *both* are important, as discussed by Celce-Murcia et al. (2010):

> Today we see signs that pronunciation instruction is moving away from the segmental/suprasegmental debate and towards a more balanced view …. Today's pronunciation curriculum thus seeks to identify the most important aspects of both the suprasegmentals and segmentals and integrate them appropriately in courses that meet the needs of any given group of learners. (2010: 11)

However, although the view that both segmental and suprasegmental features are important moves away from the segmental/suprasegmental debate, it still supports the premise that segmental and suprasegmental features can be categorized as separate entities. As will be discussed in the next section, categorizing features as either segmental or suprasegmental is not always straightforward.

Categorizing features of pronunciation: segmental or suprasegmental?

Although previous studies have tended to view segmental and suprasegmental features as separate entities, categorizing non-target-like pronunciation features as either one or the other can be problematic. Research by Zielinski (2006a, 2006b, 2008) highlighted the two-way nature of intelligibility, that is, that both the speaker and the listener play a part at times when intelligibility is reduced. She had three native listeners (native speakers of Australian English) listen to utterances produced by three L2 speakers, each from a different L1 background (Vietnamese, Mandarin, and Korean), and write down the words they heard the speakers say. At sites of reduced intelligibility (i.e., the parts of the utterances where a listener was unable to, or had difficulty in, identifying the speaker's intended words), links were made between the characteristics of the speakers' pronunciation and the listeners' difficulties identifying the words the speaker intended to say. As shown in the examples presented in Table 22.1,[2] a non-target-like feature in a speaker's pronunciation might be categorized differently depending on the perspective from which it is viewed. From the speaker's perspective we might consider both *how* a particular word was pronounced and *why* it was pronounced that way, and from the listener's perspective we might consider what the speaker was heard to say and therefore which non-target-like features were misleading.

The first example in Table 22.1 highlights a common feature of Vietnamese speakers' English pronunciation – the absence of word final consonants (see Hansen 2004). In this example, the word final consonant was absent in the speaker's production of the word *five* and the listener was misled by the absence of this consonant and heard the speaker say a non-word, *fie*. We might therefore presume that this particular breakdown in intelligibility is related to a non-target-like segmental feature (i.e., the absence of a consonant). However, when we look at the reason why the speaker might have pronounced the word in this way, we see that it is related to Vietnamese syllable structure constraints, and thus suprasegmental in nature (i.e., Vietnamese syllable structure does not allow a word final consonant following a diphthong – see Hansen 2004). As a result, it is difficult to determine whether this error should be categorized as suprasegmental or segmental.

The second example presented in Table 22.1 involves a common feature in Mandarin speakers' English pronunciation – the addition of a vowel to the end of a word (see Deterding 2006). In describing the speaker's pronunciation of the

Table 22.1 Categorizing non-target-like features of pronunciation that have an impact on intelligibility.

	The speaker		*The listeners*	
Speaker's L1	*What is the target word and how do they pronounce it?*	*Why?*	*What do they hear the speaker say?*	*What misleads them?*
1. Vietnamese	Target word: *five* Pronounced: /faɪ/ Description: final consonant /v/ is absent. → Segmental?	The syllable structure of Vietnamese does not allow a word-final consonant following a diphthong (Hansen 2004). → Suprasegmental?	*fie* (non-word)	A consonant missing at the end of the word results in them hearing a non-word. →Segmental?
2. Mandarin	Target word: *just* Pronounced: /ʤʌstə/ Description: vowel added to the end of the word. → Segmental?	Syllable structure of Mandarin does not allow for word-final consonant clusters. Adding a vowel is a common way Mandarin speakers modify them to create an open syllable structure at the end of the word (Hansen 2001). → Suprasegmental?	*just a* *just don't*	An added syllable results in them hearing an extra word. → Suprasegmental?

Notes. The Vietnamese speaker is a male from North Vietnam. The Chinese speaker is a female from Northern China. The listeners are native speakers of Australian English. See endnote 2.

word *just* in this way, the non-target-like feature (an extra vowel) would appear to be segmental in nature. However, the listeners in this example were misled by the resulting change in the syllable stress pattern (an extra vowel means an additional syllable) and both heard the speaker say two words rather than one (*just a* and *just don't*). It seems, therefore, that from the listeners' perspective, the misleading feature of this non-target-like production is suprasegmental in nature (there are more syllables than there should be). Similarly, the reason why the speaker might have produced the word this way seems to be related to the syllable structure constraints of Mandarin (see Hansen 2001), and is thus also suprasegmental in nature. Again, this raises the question of whether this feature should be categorized as suprasegmental or segmental?

As well as being difficult to do, categorization of different non-target-like pronunciation features as either segmental or suprasegmental ignores the possibility of a relationship between them, and fails to view them as part of an integrated system where one might interact with the other to influence intelligibility. The recognition of a possible interaction between the two has been raised in the speech disorder literature,[3] where Weismer and Martin (1992) argued:

> In running speech, segmental and suprasegmental events are executed simultaneously. Modifications of segmental elements … may influence not only the perception of those particular segments but also the perception of the rhythmic structure of the utterances as a whole. In this sense, the segmental event may contribute to a modification of the prosodic structure. (1992: 83)

Rather than debating whether segmental or suprasegmental features are more important, we need to rethink our approach and view the features of pronunciation as part of an integrated and interactive system where the production of one can influence the other. In this way we can further our understanding of reduced intelligibility in L2 speakers of English and gain insight into establishing not only what to teach but how to teach it in the classroom.

An integrated system of pronunciation features: the prosodic hierarchy

The prosodic hierarchy (e.g., Nespor and Vogel 1986; see Demuth 2009 for an overview) provides a useful framework for the analysis of the way different pronunciation features might combine or interact to influence a speaker's intelligibility. In this framework, the prosodic structure of spoken language is conceptualized as consisting of a hierarchy of increasingly smaller units. Within the context of the prosodic hierarchy, therefore, a particular word is seen to be composed of the units at levels below it (foot, syllable, mora) and also embedded in higher level units above (phonological phrase, intonational phrase, utterance). For example, if we use this framework to consider the word stress pattern in the word *economics* in the example presented in Table 22.2, we see that the word *economics* is composed of smaller units below and is embedded in larger units above. As described by Demuth (2009) in

Table 22.2 The prosodic hierarchy: English.

Hierarchy level	*Example*
Phonological utterance	*Some students have to study agricultural economics*
Intonational phrase	*have to study agricultural economics*
Phonological phrase	*agricultural economics*
Prosodic word	*ecoNOmics*
Foot	[NOmics] (s w)
Syllable	*mics*
Mora	*mi*

Note. The utterance used as an example here was produced by a Vietnamese speaker. See endnote 2.

English, stress at the word level is influenced by units below the word level (the shaded area in Table 22.2); that is, the mora structure tends to determine which syllables are stressed and this influences the foot structure which in turn influences stress patterns in words. Word stress patterns in English are also influenced by morphology, with different suffixes affecting where primary stress is placed in a word. Thus, the word *economics* is produced with primary stress on the third syllable.

Different languages have different prosodic constraints and therefore differ in, for example, the types of rhythmic patterns, foot structures, and syllable structures permitted. It is therefore possible that prosodic constraints of a speaker's L1 might play a role in the way they organize their pronunciation of English. Understanding what these constraints are gives us important insight into *why* speakers from particular L1 backgrounds might have the non-target-like English pronunciation features that they do, and thus provides important information about *how* we might need to go about teaching them.

The Vietnamese speaker featured in Table 22.2 actually produced the word *economics* with non-target-like stress, that is, with primary stress on the second syllable rather than the third. Consideration of the prosodic constraints of Vietnamese might therefore provide some insight into why he did so. Much of the information in the literature about why Vietnamese speakers might find English word stress challenging focuses on the transfer of different features of Vietnamese phonology. This includes observations that Vietnamese is a tonal language where most words have only one syllable (Hwa-Froelich, Hodson, and Edwards, 2002) and there seems to be no systematic difference between syllables in terms of duration or vowel quality (Nguyen and Ingram 2005) and no system of (lexical) word stress (Nguyen and Ingram 2007). It is, however, widely accepted that there is stress at the phrasal level for accentual prominence (Nguyen, Ingram, and Pensalfini 2008). If we consider these features of Vietnamese, it is no surprise that Vietnamese speakers find English word stress challenging. However, this information does not really help us understand what word stress errors Vietnamese speakers might make. We might, for example, expect that they would pronounce multisyllabic

words with equal stress on each syllable, or that perhaps they might inadvertently use tones on particular syllables, which could be perceived as stressed when they were not intended to be.

Using the prosodic hierarchy as a framework, Schiering, Bickel, and Hildebrandt (2010) have provided some insight into how Vietnamese L1 prosodic constraints might influence English word stress. They argued that, in Vietnamese, the stress pattern of a string of syllables (i.e., words) is determined by the rhythmic pattern at the phonological phrase level and is not related to a particular word (as is the case in English). They describe the rhythmic pattern at the phrasal level as sequences of up to three syllables with *ws* or *wws* patterns.

Thus, if these prosodic constraints play a role in the way a Vietnamese speaker organizes English word stress, we might expect multisyllabic words to be treated as a series of syllables that are organized into phrases according to Vietnamese constraints, and would most likely start with a *ws* or *wws* pattern, as shown in Table 22.3.

Preliminary research by Zielinski et al. (2011) supports these expectations and found that there was in fact a tendency for the Vietnamese speaker featured here to produce word stress patterns that started with *ws* rather than *sw* patterns (e.g., foREIGN [*ws*]/ foreign; ofFIcer [*wsw*]/ officer; pathoGEN [*wws*]/pathogen). They also found an instance of inconsistency, where the word *expect* was produced with target-like word stress in one context and non-target-like word stress in another. This finding suggests that the stress placed on a particular syllable in an English multisyllabic word might depend on where that word occurs in a phrase. This means that there could be some inconsistency in word stress patterns, with a particular word having target-like word stress in one phrasal context but non-target-like word stress in another. This preliminary research involved the analysis of one sample of connected speech from one speaker, and so further research is needed before any firm conclusions can be drawn. However, these findings highlight the importance of viewing different features of pronunciation as part of an integrated

Table 22.3 The prosodic hierarchy: Vietnamese.

Hierarchy level	*Example*
Phonological utterance Intonational phrase	*Some students have to study agricultural economics* *have to study agricultural economics*
Phonological phrase	*eCOnomics* *(w s) (w s)*
Prosodic word Foot	*ecoNOmics* [NOmics] (s w)
Syllable	*mics*
Mora	*mi*

system in understanding what a speaker is doing and why. They also raise questions about how we might best teach word stress patterns to learners of English from a Vietnamese L1 background, when phrasal level stress patterns might "undo" all the good work that has been done at the word level and result in a non-target-like word stress pattern being produced in a novel phrasal context. It is therefore crucial that once a particular stress pattern has been learned at the word level, the learner has the opportunity to practice this pattern in a range of different phrasal contexts, particularly those predicted to be difficult or likely to challenge the learned pattern.

By analyzing pronunciation features in the context of the hierarchical system in which they interact, we can also gain an understanding of the way in which non-target-like features at different levels might combine or interact to influence intelligibility. Zielinski (2006b, 2008), in relation to the examples presented in Table 22.1, analyzed non-target-like pronunciation features at sites of reduced intelligibility in terms of the interaction between segments (consonants and vowels), syllables (strong and weak), and words (prominent and nonprominent) within a pause group. Firstly, utterances were divided into smaller sections on the basis of the speaker's placement of pauses (thus referred to as pause groups rather than phrases). Any pause group containing a section of speech where a listener was unable to, or had difficulty in, identifying what the speaker intended to say was considered to be a site of reduced intelligibility. At each of these sites, the relative strength of each syllable was judged as either the strongest in the pause group (S), strong but not the strongest (s), or weak (w), and the segments in each syllable were identified. This analysis captured both the patterns of strong and weak syllables in multisyllabic words and the patterns of strong and weak syllables across the words in the pause group. In order to establish a link between specific non-target-like features and the words the listeners heard the speakers say, the non-target-like features replicated in the listener responses were identified. An example taken from Zielinski's study is presented in Figure 22.1 to illustrate this process.

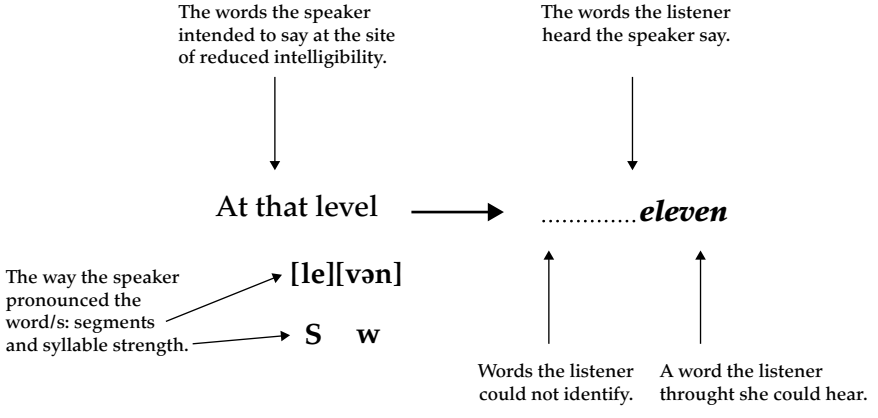


Figure 22.1 Analysis of speaker and listener contribution at sites of reduced intelligibility. Adapted from Zielinski 2006b.

The example shown in Figure 22.1 was a site of reduced intelligibility (underlined) in the following utterance produced by the Vietnamese speaker in Zielinski's study:

At that level *student have to study for five years.*

At first, the listener was unable to identify any of the words at this site, but after listening to the utterance a number of times, she commented that she thought she heard the word *eleven* in there. Thus, the speaker's syllable stress pattern and segments in the word *level* were replicated in the listener's response. In identifying this word as *eleven*, she replicated the speaker's S w stress pattern (in Australian English it is common for *eleven* to be pronounced with two syllables), but was misled by the nontarget production of the final consonant (n/l) substitution.

Such an analysis allows us to explore reduced intelligibility from the perspective of both the speaker and the listener, and investigate how non-target-like features at different levels might combine or interact to influence intelligibility. Zielinski found that, regardless of the speaker they were listening to, listeners found both the syllable stress pattern and segments produced by the speaker to be important; they all relied to some extent on both to identify the speaker's intended words. They relied consistently on the speaker's syllable stress pattern (both the number and pattern of strong and weak syllables) and more consistently on segments in strong syllables than those in weak ones. Thus, the non-target-like production of either, or both combined, had the potential to mislead the listeners and thus reduce intelligibility. Furthermore, segments in strong syllables were found to be particularly important to the listeners, especially the syllable initial consonant and the vowel. In fact, non-target-like segments in strong syllables had the greatest impact on intelligibility across all three speakers. These findings highlight the importance of moving on from the underlying assumptions inherent in the segmental/suprasegmental debate and changing our research focus to integrate segmentals and suprasegmentals.

Moving on from the segmental/suprasegmental debate

In order to move forward in our understanding of reduced intelligibility in L2 speakers of English, it is important that future research investigates how different features of pronunciation combine and interact to reduce intelligibility, and also explore the role played by both the speaker and the listener.

The interaction between different features of pronunciation

It is important that future studies analyze the speech signal in a way that allows the exploration of the way different pronunciation features might interact to influence a speaker's intelligibility. Rather than categorizing different features

as discrete items for attention, features of pronunciation need to be analyzed in the context of the integrated and interactive system of which they are part. For example, as mentioned earlier, Bent, Bradlow, and Smith (2007) investigated the relationship between intelligibility and the non-target-like production of various segments (both consonants and vowels) and found that vowel accuracy and syllable/word-initial consonant accuracy correlated with intelligibility scores. However, because syllable stress patterns were not included in the analysis, vowel changes related to the production of non-target-like syllable stress could not be distinguished from those related to non-target-like production of the vowels themselves. Similarly, Munro and Derwing (2006) investigated the impact of high and low functional load consonant errors on comprehensibility and found that high functional load errors had a greater effect than low functional load ones. Interestingly, they also noted that the high functional load errors occurred in content rather than function words. It is therefore likely that they were in strong syllables, and this may have affected the listeners' reliance on them and therefore influenced their judgments of comprehensibility. However, because only consonants were analyzed, this possibility could not be investigated.

It is also important that rather than focusing on the relative importance of individual features to intelligibility, future studies investigate the cumulative effect of multiple non-target-like features. Zielinski (2006b) found that across all three speakers in her study, it was more likely than not that multiple non-target-like features contributed to reduced intelligibility, whether it be a combination of non-target-like syllable stress patterns and non-target-like consonants and/or vowels, or a combination of different non-target-like consonants and/or vowels. Munro and Derwing (2006) investigated the cumulative effect of high and low functional load consonant errors on comprehensibility and found that neither showed cumulative effects. They speculated that the nature of the errors might be more important to comprehensibility than the number. However, they only focused on consonants, and to be able to investigate the cumulative effect of different combinations of non-target-like features, we need to analyze the speech signal in a way that allows us to consider the combination and interaction features from different levels of the integrated system in which they operate.

The role of the speaker

To further our understanding of how to improve intelligibility and comprehensibility for speakers from different L1 backgrounds, it is important that future studies explore why they have the particular non-target-like pronunciation features they do. Further investigation of the role of L1 prosodic constraints on various English pronunciation features would give us important insight into why a speaker from one L1 background might find a particular pronunciation feature challenging, while a speaker from a different L1 background might not.

The role of the listener

When listening to speech, listeners rely on speech processing strategies that are "specifically tailored" to their native language phonology (Cutler 2001: 11) and they apply these strategies regardless of who they are listening to, and whether or not it results in limited success. The native English listeners in Zielinski's studies (2006b, 2008) were misled by non-target-like features in the speakers' pronunciation because they listened to the speech with their "English ears" and relied on the non-target-like features as if they were target-like English features. Cutler describes a similar listener response to foreign language input:

> Listeners command a repertoire of procedures appropriate for their native language and not only cannot call at will upon new procedures appropriate to input in a new language but perforce apply the native procedures to the new input irrespective of whether these act to facilitate processing or to render it inefficient. (2001: 10)

As highlighted earlier in Table 22.1, because of the two-way nature of reduced intelligibility, it is important to investigate the role of the listener as well as the speaker. The listeners in Zielinski's study operated in a way that might be expected, seeing they were native speakers of English. Relying heavily and consistently on a speaker's syllable stress pattern (both the number and pattern of strong and weak syllables) is what native English listeners do, both to locate word boundaries (Cutler and Butterfield 1992; Liss et al. 1998) and for lexical access (Bond and Small 1983). The finding that they relied more consistently on segments in strong syllables than those in weak ones is also typical of native English listeners. Segments in strong syllables are important to them because they provide crucial information for lexical access (Bond and Small 1983; Bond 1999). In addition, the way segments are produced in strong syllables contributes to the perception of the syllable as strong (Cutler and Clifton 1999; Cutler 2009; Stevens 2002), and segments are less variable in strong syllables than they are in weak ones (Carroll 2004; Greenberg 2006).

An understanding of listeners' speech processing strategies is crucial to our understanding of reduced intelligibility in different contexts. The way listeners identify individual words in a stream of continuous speech is language specific and based on the listener's L1 speech processing strategies (Carroll 2004; Cutler 2001; Cutler, Dahan, and van Donselaar 1997). Listeners from different L1 backgrounds might therefore rely on different features in the speech signal to understand a speaker, and the intelligibility of the same speaker might be affected by different features for listeners from different L1 backgrounds. This poses a significant challenge to research related to the development of a Lingua Franca Core (Jenkins 2000, 2002) that would enable speakers from a wide range of L1 backgrounds to communicate effectively and intelligibly with each other. Considering the numerous combinations of speaker and listener L1 backgrounds in the mix, future research needs to not only consider what features the speaker might have in their pronunciation but also the speech processing strategies the listener might be using when listening to them.

Conclusion

The segmental/suprasegmental debate is based on a false dichotomy. Not only are both important to intelligibility and comprehensibility, but categorization of pronunciation features as either one or the other ignores the relationship between them and fails to view them as part of an integrated system where the production of one can influence the other. In order to further our understanding of reduced intelligibility and comprehensibility in L2 speakers of English, it is crucial that studies view features of pronunciation as part of an integrated and interactive system, and investigate how different features combine and interact to reduce intelligibility and comprehensibility, and why they do so for both the speaker and the listener.

NOTES

1 The terms *intelligibility* and *comprehensibility* are used here as defined by Derwing and Munro (1995). Intelligibility refers to the extent to which a speaker's utterance can be understood by a listener, and comprehensibility refers to the listener's perception of how difficult an utterance is to understand.
2 All examples presented in this chapter are drawn from data collected by Zielinski (2006b) for her PhD research. Publications related to this research are Zielinski (2006a, 2008).
3 Inclusion of a reference to the speech disorder literature does not mean that L2 speakers are considered to have a speech disorder. It was included here because it provides further information about the way different pronunciation features may influence intelligibility.

REFERENCES

Benrabah, M. 1997. Word-stress – A source of unintelligibility in English. *IRAL* 35(3): 157–165.

Bent, T., Bradlow, A.R., and Smith, B. 2007. Phonemic errors in different word positions and their effects on intelligibility of non-native speech: All's well that begins well. In: *Language Experience in Second Language Speech Learning. In Honor of James Emil Flege*, O.S. Bohn and M.J. Munro (eds.), 331–347, Amsterdam: John Benjamins Publishing Company.

Bond, Z.S. 1999. *Slips of the Ear: Errors in the Perception of Casual Conversation*, San Diego: Academic Press.

Bond, Z.S. and Small, L.H. 1983. Voicing, vowel, and stress mispronunciations in continuous speech. *Perception and Psychophysics* 34(5): 470–474.

Brown, A. 1991. Functional load and teaching pronunciation. In: *Teaching English Pronunciation. A Book of Readings*, A. Brown (ed.), London: Routledge.

Carroll, S.E. 2004. Segmentation: learning how to "hear words" in the L2 speech

stream. *Transactions of the Philological Society* 102(2): 227–254.
Catford, W.W. 1987. Phonetics and the teaching of pronunciation: a systematic description of English phonology. In: *Current Perspectives on Pronunciation: Practices Anchored in Theory*, J. Morley (ed.), 87–100, Washington, DC: TESOL.
Celce-Murcia, M., Brinton, D.M., Goodwin, J.M., and Griner, B. 2010. *Teaching Pronunciation: A Course Book and Reference Guide*, 2nd edition, New York: Cambridge University Press.
Chela-Flores, B. 2001. Pronunciation and language learning: an integrative approach. *International Review of Applied Linguistics* 39: 85–101.
Collins, B. and Mees, I.M. 2003. *Practical Phonetics and Phonology. A Resource Book for Students*, New York: Routledge.
Cutler, A. 2001. Listening to a second language through the ears of a first. *Interpreting* 5(1): 1–23.
Cutler, A. 2009. Greater sensitivity to prosodic goodness in non-native than in native listeners (L). *Journal of the Acoustical Society of America* 125(6): 3522–3525.
Cutler, A. and Butterfield, S. 1992. Rhythmic cues to speech segmentation: evidence from juncture misperception. *Journal of Memory and Language* 31: 218–236.
Cutler, A. and Clifton, C. 1999. Comprehending spoken language: a blueprint of the listener. In: *The Neurocognition of Language*, C.M. Brown and P. Hagoort (eds.), 123–166, Oxford: Oxford University Press.
Cutler, A., Dahan, D., and van Donselaar, W. 1997. Prosody in the comprehension of spoken language: a literature review. *Language and Speech* 40(2): 141–201.
Demuth, K. 2009. The prosody of syllables, words and morphemes. In: *Cambridge Handbook on Child Language*, E. Bavin (ed.), 183–198, Cambridge: Cambridge University Press.
Derwing, T.M. and Munro, M.J. 1997. Accent, intelligibility, and comprehensibility. Evidence from four L1s. *Studies in Second Language Acquisition* 19: 1–16.
Derwing, T.M. and Munro, M.J. 2005. Second language accent and pronunciation teaching: a research-based approach. *TESOL Quarterly* 39(3): 379–397.
Derwing, T.M. and Munro, M.J. 2009. Putting accent in its place: rethinking obstacles to communication. *Language Teaching* 42(4): 476–490.
Derwing, T.M., Munro, M.J., and Wiebe, G. 1997. Pronunciation instruction for "fossilized" learners. Can it help? *Applied Language Learning* 8(2): 217–235.
Derwing, T.M., Munro, M.J., and Wiebe, G. 1998. Evidence in favor of a broad framework for pronunciation instruction. *Language Learning* 48(3): 393–410.
Deterding, D. 2006. The pronunciation of English by speakers from China. *English World Wide*, 27(2): 175–198.
Fraser, H. 2001. *Teaching Pronunciation: A Handbook for Teachers and Trainers. Three frameworks for an integrated approach*, NSW: Department of Education Training and Youth Affairs.
Gilner, L. and Morales, F. 2010. Functional load: transcription and analysis of the 10,000 most frequent words in spoken English. *The Buckingham Journal of Language and Linguistics* 3: 135–162.
Greenberg, S. 2006. A multi-tier theoretical framework for understanding spoken language. In: *Listening to Speech: An Auditory Perspective*, S. Greenberg and W.A. Ainsworth (eds.), 411–433, Mahwah, NJ: Lawrence Erlbaum Associates.
Hahn, L.D. 2004. Primary stress and intelligibility: research to motivate the teaching of suprasegmentals. *TESOL Quarterly* 38(2): 201–223.
Hansen, J.G. 2001. Linguistic constraints on the acquisition of English syllable codas by native speakers of Mandarin Chinese. *Applied Linguistics* 22(3): 338–365.

Hansen, J.G. 2004. Developmental sequences in the acquisition of English L2 syllable codas. *Studies in Second Language Acquisition* 26: 85–124.

Hwa-Froelich, D., Hodson, B.W., and Edwards, H.T. 2002. Characteristics of Vietnamese phonology. *American Journal of Speech-Language Pathology* 11(3): 264–273.

Isaacs, T. and Trofimovich, P. 2012. Deconstructing comprehensibility. *Studies in Second Language Acquisition* 34: 520–531.

Jenkins, J. 2000. *The Phonology of English as an International Language: New Models, New Norms, New Goals*, Oxford: Oxford University Press.

Jenkins, J. 2002. A sociolinguistically based, empirically researched pronunciation syllabus for English as an International Language. *Applied Linguistics* 23(1): 83–103.

Kang, O. 2010. Salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System* 38(2): 301–315.

Levis, J. 2005. Changing contexts and shifting paradigms in pronunciation. *TESOL Quarterly* 39(3): 369–377.

Liss, J.M., Spitzer, S., Caviness, J.N., Adler, C., and Edwards, B. 1998. Syllable strength and lexical boundary decisions in the perception of hypokinetic dysarthric speech. *Journal of the Acoustical Society of America* 104(4): 2457–2466.

Munro, M.J. and Derwing, T.M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning* 45(1): 73–97.

Munro, M.J. and Derwing, T.M. 2006. The functional load principle in ESL pronunciation instruction: an exploratory study. *System* 34: 520–531.

Nespor, M. and Vogel, I. 1986. *Prosodic Phonology*, Dordrecht: Foris Publications.

Nguyen, T.A.T. and Ingram, J. 2005. Vietnamese acquisition of English word stress. *TESOL Quarterly* 39(2): 309–319.

Nguyen, T.A.T. and Ingram, J. 2007. Acoustic and perceptual cues for compound-phrasal contrasts in Vietnamese. *Journal of the Acoustical Society of America* 122(3): 1746–1757.

Nguyen, T.A.T., Ingram, J. and Pensalfini, J.R. 2008. Prosodic transfer in Vietnamese acquisition of English contrastive stress patterns. *Journal of Phonetics* 36(1): 158–190.

Rogers, C.L. and Dalby, J. 2005. Forced-choice analysis of segmental production by Chinese-accented English speakers. *Journal of Speech, Language, and Hearing Research* 48: 306–322.

Saito, K. 2011. Examining the role of explicit phonetic instruction in native-like and comprehensible pronunciation development: an instructed SLA approach to L2 phonology. *Language Awareness* 20(1): 45–59.

Schiering, R., Bickel, B., and Hildebrandt, K.A. 2010. The prosodic word is not universal, but emergent. *Journal of Linguistics* 46: 657–709.

Stevens, K.N. 2002. Toward a model for lexical access based on acoustic landmarks and distinctive features. *Journal of the Acoustical Society of America* 111(4): 1872–1891.

Tanner, M.W. and Landon, M.M. 2009. The effects of computer-assisted pronunciation readings on ESL learners' use of pausing, stress, intonation, and overall comprehensibility. *Language Learning and Technology* 13(3): 51–56.

Weismer, G. and Martin, R.E. 1992 Acoustic and perceptual approaches to the study of intelligibility. In: *Intelligibility in Speech Disorders*, R.D. Kent (ed.), 67–118, Amsterdam: John Benjamins Publishing Company.

Zielinski, B. 2006a. The intelligibility cocktail: an interaction between speaker and listener ingredients. *Prospect: An Australian Journal of TESOL* 21(1): 22–45.

Zielinski, B.W. 2006b. Reduced intelligibility in L2 speakers of English. PhD, La Trobe University, Bundoora.

Retrieved from:http://arrow.latrobe.edu.au:8080/vital/access/manager/Repository.

Zielinski, B.W. 2008. The listener: no longer the silent partner in reduced intelligibility. *System* 36(1): 69–84.

Zielinski, B., Yuen, I., Demuth, K., and Yates, L. 2011. Rethinking theory for practice in pronunciation teaching. Paper presented at the American Association of Applied Linguistics, Chicago.

23 Applying Theories of Language and Learning to Teaching Pronunciation

GRAEME COUPER

Introduction

This chapter begins by considering the role of theory and its relevance and usefulness in the classroom. This leads to a review of the relevance of several theoretical positions across disciplines:

1. Applied Linguistics: Second Language Acquisition (SLA) theory;
2. Educational Psychology: social theories of learning;
3. Phonology and L2 Speech Research; and
4. Cognitive Linguistics and Phonology: a pronunciation learning and teaching framework.

The second half of the chapter describes how insights from theory can be translated into practice. These approaches and techniques, supported by research that has found them to be of value, are presented as a series of tips for teachers:

1. Teaching tip one: understand all is not as it seems;
2. Teaching tip two: generate dialogue;
3. Teaching tip three: establish category boundaries through Critical Listening;
4. Teaching tip four: meaningfully integrate pronunciation into further practice activities;
5. Teaching tip five: provide the right kind of corrective feedback.

The Handbook of English Pronunciation, First Edition. Edited by Marnie Reed and John M. Levis.

What can theories tell us and which ones should we listen to?

Jordan (2004) suggests that attempting to explain phenomena is fundamental to theory building. However, in SLA theory there is still a lack of agreement as to what the phenomenon of language actually is. At one end of the spectrum Gregg (2001), for example, puts forward the traditional SLA view that it is a matter of linguistic competence or knowledge of language inside the brain that counts. In contrast to this, other theorists suggest that what is of interest is the way language is used to communicate. This focus on meaning as opposed to rules may be broadly referred to as a usage-based approach.

Further, many theorists see little connection between SLA theory and L2 instruction. Most provocatively, Gregg (2001: 153) suggests that those who insist on a connection should "get the hell out of the armchair" and goes on to state that "SLA still hasn't shown any theoretically relevant relation between some specific type of input modification on the one hand, and some specific bit of acquisition on the other" (2001: 169). Others such as van Patten (2010) frame the gap between theory and practice somewhat differently, suggesting that SLA can help teachers to understand the acquisition process that may inform their teaching. He is, however, rather pessimistic about teachers being able to apply this as he considers they are subject to pedagogic grammars, which consist of rule learning. The aim of this chapter is to show that some theory does indeed have useful classroom applications.

Theories from a wide range of disciplines are examined to determine what they may offer in the way of help to us as teachers in the classroom. We begin with SLA theory, reviewing the range of views from the nativist through to the skills-based, before moving on to look at what guidance can be found from educational psychology and related social theories of learning. These have received more attention since the social turn in SLA (Block 2003), which moves from an acquisition to a participation metaphor (Ortega 2011), and they offer hope of practical guidance for the classroom. Our focus then moves to different understandings of phonology, from generative (Chomsky and Halle 1968) to usage-based (Bybee 2001), and L2 speech research. L2 speech research has investigated the extent to which adults can learn L2 phonological categories. This focus on learning how to conceive new categories then leads to Cognitive Phonology, within Cognitive Linguistics, which provides a useful framework for bringing together both the cognitive and social aspects of pronunciation learning.

SLA theory

SLA theory has devoted a great deal of attention to the role of cognition, dealing with issues such as whether or not there is a separate language acquisition device (LAD) or the possibility of an interface between explicit and implicit knowledge (R. Ellis 2009). Figure 23.1 is an attempt to represent the spectrum of views on the role of

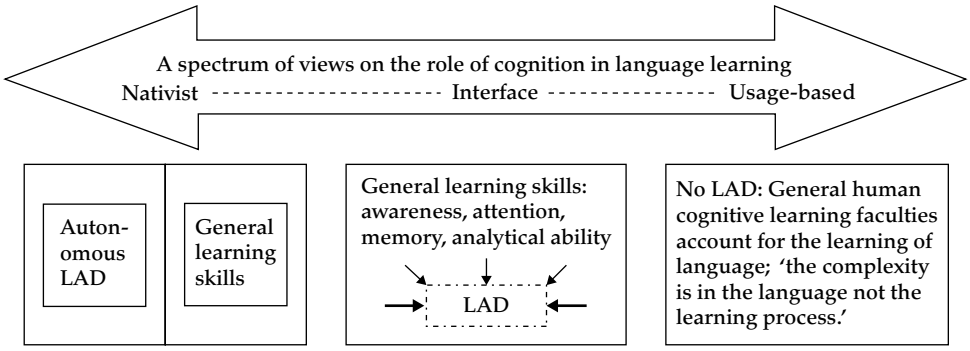


Figure 23.1 A spectrum of views on the role of cognition in language learning.

cognition in language learning. These range from positions that rely on an innate language acquisition device (LAD) to usage-based views, or what N. Ellis (2001: 37) refers to as a constructivist approach, which denies innate linguistic universals and claims that "the complexity is in the language, not the learning process".

The nativist model was the basis of SLA up until the end of the 1980s (Macaro 2003) and was behind Krashen's (1982) notions of comprehensible input and the monitor. Krashen proposed a strict distinction between uninstructed "acquisition" and instructed "learning", leading to implicit and explicit knowledge respectively. This is often referred to as the noninterface position "that learned knowledge can never become acquired knowledge" (Doughty 2003: 258). This reduces the teacher's role to supplying comprehensible input as instruction would not be not expected to have any effect on "acquisition" (Doughty 2003; Housen and Pierrard 2005).

Consideration of broader aspects of language learning, such as the role of perception and memory, has led many SLA theorists to conclude that there may in fact be some sort of interface allowing explicit knowledge to become implicit (Doughty 2003) and by implication that SLA is open to instruction (Housen and Pierrard 2005). There is a range of views as to how strong this interface is. R. Ellis (2006, 2009) takes a weak interface position, suggesting that explicit knowledge can in some way trigger the acquisition of implicit knowledge. There are, however, difficulties in defining and measuring explicit and implicit knowledge and learning (Ellis 2009). For example, explicit knowledge could refer to knowing the metalanguage, but it could also, and more usefully, refer to being consciously aware of how a structural feature works. Thus, different types of explicit knowledge may have different implications for the learning process. A distinction is also often drawn between declarative knowledge, knowing "that", and procedural knowledge, knowing "how" (e.g., Macaro 2003). These views also imply some role for explicit instruction.

Those who take a usage-based approach to language deny the role of the LAD and suggest that general human cognitive learning faculties account for the learning of language, i.e., it can be viewed as learning a skill, as proposed by Anderson (1993) and DeKeyser (1998).

Those in SLA who allow for the role of explicit instruction see the importance of attention, noticing, and awareness in learning. Schmidt (2001: 3–4) suggests that "the concept of attention is necessary in order to understand virtually every aspect of SLA". He goes on to hypothesise that "SLA is largely driven by what learners pay attention to and notice in target language input and what they understand the significance of noticed input to be". Attention is often seen as the mechanism that controls awareness, suggesting that the level of attention may lead to different levels of awareness: perception, detection, noticing, or understanding (Housen and Pierrard 2005). Schmidt defines noticing as attending to surface features, or instances of language, as opposed to metalinguistic awareness of abstract rules or principles. Noticing involves learners in paying attention to the gap between their production and the target. Thus it is an important step in the process of acquisition. This clearly has useful implications for the classroom.

SLA has also focused on the potential role of corrective feedback (CF). While a noninterface position places no value on CF, Schmidt's noticing hypothesis clearly allows for it and Long's (1996) interaction hypothesis suggests that the negotiation of meaning and recasts can be a useful source of CF. Swain's (1995) output hypothesis claims a role for CF in modified output leading to language learning.

While explicit instruction has been found to be effective, SLA theory does not provide distinctions between different types of explicit instruction (Norris and Ortega 2001) and nor does it help to identify the variables that might make one teaching technique more effective than another (Ellis 2002: 50). Housen and Pierrard (2005: 11) also note that some features are more open to instruction than others and also that "Metalinguistic rules and pedagogical descriptions can differ in clarity, intelligibility and processability so that a given target feature can be explained in both simple and elaborated terms." Widdowson (2003: 111) discusses the differences between linguistic and pedagogic descriptions and notes that "different descriptions focus on different aspects of the truth". This raises the question of how learners perceive both the evidence they are presented with and accompanying explanations, and leads us on to a consideration of social theories of learning.

Social theory and educational psychology

While the cognitivist views described above have mainstream acceptance within SLA, calls to acknowledge the social nature of language learning have led to a social turn in SLA (Block 2003; Ortega 2011). This movement towards a recognition of the embedded nature of learning allowing for the inclusion of a sociocultural perspective (Zuengler and Miller 2006) is often related to a participation metaphor as opposed to the acquisition metaphor used to describe much mainstream SLA (Ortega 2011). Indeed, there have been many calls for a greater focus on the social aspects of language learning (e.g., Atkinson 2011; Block 2003; Lantolf 1996; Lantolf and Thorne 2006) as there has been increased recognition that knowledge is formed through interaction with a social context (Sanz 2005).

A number of theories within the field of educational psychology are of particular interest to language teachers, e.g., Socio-Cultural Theory (SCT), Social Interactionism,

and Constructivism. SCT takes into account the context of communication and views it as "emergent and cumulative based on shared knowledge with an interlocutor, and that communication involves the sending and receiving-constructing of 'assumptions' (rather than stand-alone 'messages')" (Thorne 2000: 228). Thus it views "language use in real-world situations as fundamental, not ancillary, to learning" (Zuengler and Miller 2006: 37). Lantolf and Thorne (2006: 4), in describing the role of SCT in language, argue that "because SCT is a theory of mediated mental development, it is most compatible with theories of language that focus on communication, cognition and meaning rather than on formalist positions that privilege structure". Here "meaning" refers to conceptual meanings that mediate thinking rather than referential meaning. Vygotsky's SCT describes how learners gain control and independence when they "appropriate mediational means, such as language, made available as they interact in socioculturally meaningful activities" (Zuengler and Miller 2006: 39). This suggests both teachers and peers have a role to play in the language development of learners.

Williams and Burden (1997) review influences on language teaching from different approaches to educational psychology. They note that constructivists, unlike behaviourists, understand that "the sense that learners themselves seek to make of their worlds, and the cognitive or mental processes that they bring to the task of learning" (1997: 12) are essential parts of the learning process. Therefore the teacher and student can co-construct meaning as they bring their subjective realities together, mediating learning and the formation of new concepts. Macaro (2003) notes the relevance of this approach in language teaching and Blyth (1997: 51) suggests its usefulness when dealing with poor textbook explanations, which present rules as if they were a "direct reflection of an objective reality".

A further theoretical approach to come from educational psychology is social interactionism which maintains "we learn a language through using the language to interact meaningfully with other people" (Williams and Burden 1997: 39). Williams and Burden propose social interactionism, encompassing both cognitive and humanistic perspectives as "a much-needed theoretical underpinning to a communicative approach to language teaching" (1997: 39). This framework emphasizes the dynamic interaction between teacher, learner, task, and context. In the second language learning context the teacher and the learner must interact to establish meaning through effective cross-cultural communication.

Other theories that have focussed on the socially and situationally embedded nature of language learning include Atkinson's sociocognitive approach (2011), which focuses on the integrated role of the mind, body, and world in SLA. This holistic approach is also represented in Acton's use of haptics based on movement and touch (2013).

The idea that we co-construct discourse in the classroom underlies Swain's (2000: 112) extension of her output model "to include its operation as a socially-constructed cognitive tool. As a tool, dialogue serves second language learning by mediating its own construction, and the construction of knowledge about itself". More recently Swain (2006) has introduced the term "languaging" to describe this. Gibbons (2006) also analyses the bridging role of talk between teachers and

students as co-constructed discourse. The theory of Intercultural Language Teaching (Lo Bianco, Liddicoat, and Crozet 1999) is also based on the importance of cross-cultural communication occurring in a third place where conceptual and experiential learning and conceptual learning can take place with the help of dialogue in developing "a shared, pragmatic understanding of what we're talking about" (Carr 1999: 105). This supports the role of explicit instruction "using a new metalanguage which enables both teachers and learners to talk about language and culture" (Crozet and Liddicoat 1999: 121).

Phonology and L2 speech research

Having reviewed some of the theories related to language learning we will now look at theories that focus specifically on phonology and the learning and teaching of pronunciation.

Phonology involves abstract categories such as sets of segments (phonemes), tones, intonation, and voice quality (Shockey 2003). The fact that these categories are abstract is generally acknowledged, e.g., Fromkin et al. (1996). However, generative and usage-based views of phonology are in disagreement as to the nature of those categories.

Generative phonology views the categories of phonology as being determined by the presence or absence of certain distinctive features (Chomsky and Halle 1968). That is, it assumes underlying phonological rules that can be acquired through access to a Universal Grammar (UG). This implies that the teacher should focus on the physical production of sounds, i.e., the motor skills, because:

- these rules are innate and cannot be taught (Krashen 1982) and
- speech is "no more than the transmission phase of language" (Cruttenden 2001: 296), i.e., it is seen as the physical representation of language but is somehow separable from the underlying meaning.

However, as Chomsky and Halle (1968: 3) note, pronunciation is not just a matter of phonological rules, there are "many other factors as well – factors such as memory restrictions, inattention, distraction, nonlinguistic knowledge and beliefs, and so on". Clearly the impact of "other factors" on performance is much greater for the L2 learner, which leads us to look for different theoretical positions with greater explanatory power in terms of language learning and teaching.

Bybee (2001: 34) proposes a usage-based model that "goes beyond structuralist models to show how language use gives rise to structure". This model views pronunciation as an integral part of the meaning-making process rather than the transfer of a set of underlying phonological rules as it observes how speakers categorize language and how they relate the physical sounds to meaning. Bybee concludes that phonological categories are based on exemplars and the development of prototypes. The value of a usage-based approach is that it focuses on meaning and not on rules that create the impression of dichotomous features such as the voiced/voiceless distinction, when in fact they overlap (Mompean 2014). The

implications for the teacher and learner are that it is possible through cognitive skill learning processes to help learners understand the relationship between sound and meaning. The difficulties in understanding this relationship are explored in the first tip for teachers, where we consider the implications of the gap between phonology and the physical sounds of speech.

L2 speech research helps in understanding speech perception and provides us with a number of insights into category formation processes. Kuhl and Iverson's (1995) perceptual magnet effect suggests that sounds are perceived in terms of the prototype categories for the language or languages we know. We then assimilate nonprototypical members into the prototype and shrink the acoustic-phonetic space towards it. Therefore "L1 prototypes constrain learners' abilities to perceive contrasts in L2 by the 'pulls' they exert" (Leather 1999: 5). Flege (1990: 255), in describing what is required for the process of speech learning to be successful, notes that one must be able to "establish central perceptual representations for a range of physically different phones ('sounds') which signal differences in meaning, and [develop] motoric routines for outputting sounds in speech production". As well as a prototype for each sound category, there is evidence for episodic effects, that is, one remembers the particular examples one has heard and these are called upon when categorizing sounds (Pisoni and Lively 1995).

Flege's (1995) Speech Learning Model (SLM) provides useful insights into L2 pronunciation learning. It "aims to account for age-related limits on the ability to produce vowels and consonants in a native-like fashion" (1995: 237). It is assumed that "our phonetic systems remain adaptive over the life span and reorganise to allow for L2 sounds by adding new phonetic categories or modifying old ones" (1995: 233). To do this we must discern some of the phonetic differences and be able to relate the L1–L2 sounds at an allophonic level. With age we may find it harder to notice subtle but possibly significant differences and classify similar L1–L2 sounds as being the same. This model allows for the influence of the L1 and for the impact of an L2 on modifying existing categories and forming new ones. Therefore it directs our attention to the processes of category formation and the role that training may play.

While the SLM focuses on the formation of categories, another approach is to focus on how the stream of sound can be interpreted by the listener to recognize words and the phonemes that make them up psychologically. Lively, Pisoni, and Goldinger (1994: 265) explain that this is complex for a number of reasons. Firstly, there is a lack of acoustic–phonetic invariance, i.e., acoustic forms of words and phonemes are different when produced by different speakers, but they are also different when produced by the same speaker in different occurrences, in different situations, and phonetic contexts. Secondly, phonemes are not produced linearly; they overlap and are co-articulated, making it impossible to reliably map acoustic features to perceived phonemes. Finally, there is a lack of segmentation, which means we have to rely on context-sensitive cues such as stress and intonation contours to assist.

One of the major difficulties in learning a new category is to discern those aspects of the auditory signal that mark the sound as belonging to that particular

category. This may lead to learners attending to a difference that is of little significance to the native speaker. A number of researchers have noted how difficult it can be to get learners to focus on the "right" cues (e.g., Munro and Bohn 2007). However, Guion and Pederson (2007) investigated the role of attention in phonetic learning and concluded that explicit directing of attention can help adult learners to discern novel phonetic contrasts better.

Cognitive Linguistics and Phonology: a pronunciation learning and teaching framework

In reviewing theories that have the greatest explanatory power from the perspective of what goes on in the classroom, I have found Cognitive Phonology (CP), derived from Cognitive Linguistics, provides a coherent explanation for the phenomenon of adult L2 pronunciation learning. CP's conclusions, although not always arrived at in the same manner, are congruent with those of many of the theories reviewed above: pronunciation is a cognitive skill that can be learned using our general learning faculties as proposed by the interface position from SLA theory; pronunciation learning is situationally embedded involving a complex interplay of social and cognitive variables in the construction of meaning as suggested by SCT and other socially oriented theories that adopt a participant metaphor. Lantolf (2011) notes that SCT fits in comfortably with Cognitive Linguistics; CP also takes a usage-based approach to phonology and is in line with many of the findings from L2 speech research.

Cognitive Phonology is a branch of Cognitive Grammar, situated within Cognitive Linguistics, stemming largely from the work of Langacker (1987) and Taylor (2002). Cognitive Linguistics is based on the premise that the cognitive abilities required for language are similar to those required for other learning and "it argues that language is embodied and situated in the sense that it is embedded in the experiences and environments of its users" (Mompean 2006: vii). It uses what is known about cognition to build theories of language acquisition rather than the other way around and it totally rejects the Chomskyan view in Generative Theory that language is in the mind and autonomous (Taylor 2002). It also rejects computer processing type analogies for the way the brain processes language and is distinct from cognitive psychology, which focuses more on subconscious processing on inaccessible mental representations (Anderson 2000).

Mompean (2014) reviews the main implications of Cognitive Linguistics for the understanding of phonology in terms of two guiding assumptions. The first is "language, including phonology, is the outcome of properties of cognition" (2014: 357). He analyses the importance of three cognitive abilities with relevance for phonology: categorization, perception, and conceptual combination. Categories are considered central to conceptual and linguistic organization but most notably Cognitive Linguistics does not accept the traditional view of categories as being discrete and defined by necessary and sufficient conditions with features distributed evenly across category members. Rather, categories are defined through "overlapping similarities with different category members or similarity to a central

or prototype member of the category" (2014: 360). Perception is also a cognitive ability. It enables us to recognize similarity, which leads to categorization. Another relevant perceptual capacity is attention to salience, which enables us to distinguish between the figure, or what needs to be heard as prominent, and ground, which does not require our attention. The implications of this for teachers and learners are taken up in the next section.

The second assumption of Cognitive Linguistics is "linguistic organisation (phonological inclusive) is also the outcome of the bodies humans have and how they interact with the sociophysical world" (Mompean 2014: 357). This focus on the social and cultural aspects of language and language learning is in line with the social turn seen within SLA.

Teaching tips

Fraser (2006, 2010), noting that categories, concepts, and concept formation are central to CP, has applied this theory to pronunciation teaching and learning. This begins with the understanding that it is the concepts, or mental representations of categories, that allow us to categorize (Murphy 2002). Because phonological concepts are language-specific, when we learn a new language we have to learn how the speakers of that language conceptualize, or think about, its categories. Couper (2006, 2011, 2012, 2013) has undertaken a series of classroom-based studies, which have investigated how we as teachers can help learners to form these concepts in order to accurately categorize the sounds of the new phonological system. The practical implications of this research are explained in the following tips for teachers.

Teaching tip one: understand all is not as it seems

Applying theory from Cognitive Phonology to the classroom situation, the first thing we need to remember is that our phonological concepts determine how we categorize sounds and that these concepts are language-specific. That is, we perceive speech differently so when learning L2 pronunciation we have to learn a different way of thinking about sounds. As teachers we have to remember that when we think about English pronunciation we are thinking about it through a filter built up through many years, maybe even a lifetime, of experience in extracting meaning by categorizing sound into relevant categories. We are so proficient at it that it is easy to forget that the actual sounds produced do not relate one-to-one with the phonemes we see in the dictionary. So there is a two-step process: first we have to understand the difference between what we actually say (the physical sounds) and what we think we say (the phonology, or the way we categorize sounds). Then we have to help students to go through the same process with their first language by getting them to use their ears to move away from the way they are used to thinking about sounds so that we can help them to understand the way English speakers think about sounds.

As Shockey (2003: 10) notes, "most people speaking their native language do not notice either the sounds that they produce or the sounds that they hear". Shockey also reports that often whole sounds are omitted even though the listener still perceives them. Phonologically it is easy to think of pronunciation as a sequence of one sound after the other. However, acoustically it has been demonstrated that this is not in fact the case. Warren (1982) provides experimental evidence to show that speech perception is not dependent on an ability to identify component sounds and their orders, that in fact a great deal of speech would be too fast to do this, and that we rely on holistic pattern recognition.

Fraser (2004), taking the example of the words *bat* and *bad*, provides a good demonstration of the difference between what we think we hear and the physical sounds and the implications of this for teachers. While most naïve listeners will say the difference between *bat* and *bad* is that one has a /t/ and the other has a /d/, in fact the greatest difference is in the length of the vowel. Acoustically the /d/ and /t/ are surprisingly similar. Of course this understanding of phonemes is also often supported by spelling, which misleads both teachers and learners through what is known as a literacy bias (Linell 2005). This is one simple example of the difference between what we perceive and the actual sounds, and one can see the significance of this in a teaching context. If a teacher insists the learners produce a /d/ or a /t/ we are likely to see the unexpected production of aspiration or an additional vowel (referred to as epenthesis) at the ends of words.

An analogy of how speech perception works can be seen through visual perception puzzles. Take, for example, the picture in Figure 23.2. When you look at the picture, you first need to understand that there is another way of looking at it. You might see a young lady. I can explain this by saying I see an old lady. However, this might not help you to see it, it just tells you to keep looking. I can try and explain what to look for, to see the old woman. She has a big nose. She's looking down. Her mouth is a thick black line. If this doesn't help, I could put the salient lines into the foreground and push the other ones into the background by drawing the outline of the old woman's nose and eye (as in Figure 23.3). I may also have to think of other ways to make the second perspective clearer to you. Now you should be able to see what the differences are. Of course when you look again later, you may still have difficulty in finding the second perspective.

This is an example of how figure ground organization works, which demonstrates the sorts of difficulties our students might be having when trying to adjust their figure ground perception. As was noted in the review of L2 speech research, it can be difficult to get students to notice the salient cues. Cognitive Phonology provides the explanation that as in different languages different aspects of incoming acoustic data are phonologically salient, one has to learn what is salient in the target language in order to form the concepts required for the L2's phonological categories. As teachers, it is important to understand this if we are going to be effective in helping learners to form new concepts.

Therefore the first tip is to make sure you can move away from your subjective perception of L1 speech. It does not feel subjective because that is how your speech

Figure 23.2 Visual perception. For many years, the creator of this figure was thought to be British cartoonist W.E. Hill, who published it in 1915 in *Puck* humor magazine, an American magazine inspired by the British magazine *Punch*. However, Hill almost certainly adapted the figure from an original concept that was popular throughout the world on trading and puzzle cards.

Figure 23.3 Adding lines for a different view.

community perceives it, which in turn makes it feel like reality, or the truth. Therefore we need to untruth, or step outside, our normal view of the world. If we do not do this and instead hold to our perceptions as being the objective truth of what those sounds are, we are in danger of insisting to our students that they need to produce and hear certain sounds that are not in fact there. If we can recognize the physical sounds and the gap between them and the perceived sounds, we are in a better position to help our students make the connection between form and meaning. While the more teachers understand about phonology the better, the main point is to learn to be able to step aside from your own perceptions, or to untruth, and listen to sounds more neutrally to try and imagine how they might sound to your students.

Teaching tip two: generate dialogue

Given the different phonological perceptions of teachers and students, effective communication about pronunciation requires that we establish common understandings with our students. For example, when talking to students about syllables it is easy to think of it as a category and forget that the concept behind this category varies from language to language. This leads to the situation where we fail to communicate with our students because on the one hand English conceives of a syllable as containing at least one vowel, which may have a number of consonants in the onset and coda. On the other hand, other languages have different concepts of syllables, such as consisting of one consonant followed by one vowel. We may not be consciously aware of these concepts but they control how we perceive speech. These cross-linguistic differences in concepts may mean that if my L1 follows a consonant vowel syllable structure I am likely to relate the syllable to the presence of a consonant, so that every time I come across a consonant I will expect a vowel to follow. They may also render talking to students about syllables useless (Couper 2006) unless we find ways to communicate about them. This involves the development of a common way of talking, or the social construction of metalanguage. I observed the value of co-constructing such a language through a number of classroom interactions and found empirical evidence for its effectiveness in improving pronunciation (Couper 2011). I refer to this language as socially constructed metalanguage (SCM).

Socially constructed metalanguage (SCM) refers to the kind of metalanguage that is needed for effective metalinguistic communication. Such communication, as with all cross-cultural communication, relies on both parties having a common understanding of the concepts that are being discussed. SCM requires the teacher and the learners to work together to construct common ways of talking about these concepts. This involves the teacher in understanding how the learners interpret the sounds of the target language. One way the teacher can do this is by asking learners to describe the difference between two productions. Equally it involves the learners in understanding how the sounds they produce are interpreted by the native speaker. It is social in the sense that it is owned by the class as a group and

it refers to the social nature of language learning and the role of social construction of meaning. Once this metalanguage has been developed, it can be used throughout the course for quick and effective feedback. While the term SCM has been developed in relation to teaching pronunciation, it could just as easily apply to the use of explicit instruction in all aspects of language teaching.

In practice, this means that we need to start by asking our students to tell us how they understand the sounds of English. For example, if a student says fishy when they want to say fish, I write the two words on the board. I will explain that, to my ears, it sounded like fishy, pointing to and underlining the difference. I can then model the two words, asking the student to tell me how they are different. They are unlikely to say there is an extra syllable; rather, they will suggest the "shy" in fishy is longer or louder. Alternatively, they might suggest the "sh" in fish is shorter or quieter. This tells me that while I perceive an extra syllable, they simply perceive it as a different way of saying the same sound. This means we need to help the learners understand the salient differences between the two for the proficient English speaker, in other words to establish the phonological category.

To do this, I will ask the student to say both words and I will point to the one I hear. In giving them feedback, I can use the language they have already used to describe the differences between the sounds. So I might tell them to make the "sh" shorter or quieter to help them produce fish rather than fishy. Of course they still need a great deal of practice, especially if they have been saying it incorrectly for a long time. However, once the learner understands how these two sounds are categorized differently by English speakers they can remind themselves what they have to do to get the message across.

By engaging in this sort of dialogue with our students we can focus on all aspects of pronunciation. For example, a recent study (Couper 2012) focused on word stress, beginning with learners' current perceptions of word stress in both their own languages and English and helping them to understand how the concept of stress was actually different in different languages. Again, by having this classroom conversation, common understandings were developed and communication was more effective when providing explanations and feedback. What it amounts to is effective cross-cultural communication, enabling teachers and learners to better understand each other and develop a common basis on which to build language proficiency.

While this approach was developed directly from Cognitive Phonology, theories that focus on language awareness and social theories such as SCT might also support this approach. What it offers over the traditional SLA approach is that it defines a specific type of explicit instruction that it suggests is better than any old type of explanation. Indeed, Couper (2011) provides convincing empirical evidence to support this claim. With regards to the need for explicit as opposed to implicit instruction, some of us may learn many of these L2 phonological concepts implicitly, but there will usually be some concepts that require explicit intervention. This is where the teacher needs to be aware of the need to provide this sort of instruction.

Teaching tip three: establish category boundaries through Critical Listening practice

As we learn L2 phonological concepts we are learning how sounds are categorized in the L2. As a part of this process we need to establish the prototypical sound as well as all the variations that would be classified as belonging to the same category. Fraser (2009) suggests Critical Listening as a way of helping learners in this process. Critical Listening involves the learner in listening for the contrast between two productions: one that is acceptable and one that is not. Typically there should be a meaningful difference, and ideally it would involve comparing the learner's production when it is acceptable with when it is not. As with SCM, it involves helping learners to understand how the sounds are perceived by the native speaker. It involves a focus on developing speech perception and learning where the boundaries are between the different phonological categories. Again, this approach is derived from Cognitive Phonology.

In practice this might involve learners recording themselves and then listening to their recording and comparing it with a model in conjunction with getting feedback from peers or the teacher. Even though this has the potential to be face threatening when done with a large class, I have found that as long as you can develop the appropriate supportive atmosphere the students will work cooperatively to help establish these category boundaries. This helps not just the particular learner but all other learners as well as they can learn more examples of what belongs to these categories. It is important to note that these exercises should focus on meaningful differences rather than what might be construed as slight differences in accent. The question is whether the target language speaker interprets the sounds as intended.

Fraser (2009: 301) provides evidence to support the claim that Critical Listening, focusing "on the contrast between a correct (or appropriate) pronunciation versus an incorrect (or inappropriate) pronunciation within a particular communicative act", could help in forming new phonological concepts, in this case establishing the boundaries between the /r/ and /l/ categories in English. Further support was found by Couper (2011) in teaching learners to produce syllable codas without epenthesis, or producing an additional vowel.

This technique is also supported by findings from L2 speech research, which has clearly shown that adults can be trained through comparing and contrasting to learn these categories and their boundaries. Strange (1995: 40) reviews research into the effect of training and concludes adults can learn new phonological contrasts as they "retain the auditory perceptual abilities that are required for the detection and discrimination of the acoustic parameters that carry phonetically relevant information", i.e., the right kind of training can help adult learners to improve their L2 speech perception. Rochet (1995) also concludes that difficulties adults have in perceiving L1–L2 differences in similar phonemes is not representative of a sensory-based loss but rather of a change in selective attention.

One way to help set up Critical Listening practice is to get students to record themselves at the beginning of a course and use this as a diagnostic giving them

initial feedback and encouraging them to set up their own goals for improving their pronunciation. The teacher can use these recordings to prepare examples for Critical Listening that contrast different productions of the same word or utterance.

Teaching tip four: meaningfully integrate pronunciation into further practice activities

Cognitive Linguistics assumes that form is motivated by meaning and the way we relate meaning to form is determined by our perception (Holme 2012). We have already seen how SCM and Critical Listening are important in the formation of L2 phonological concepts. However, to fully form and automatize these concepts we often need further practice. While SCM and Critical Listening involve meaningful and genuine communication about language and the relationship between form and meaning, we also need to help learners use this language in communicative situations. Here, teachers can draw on their experience with communicative language teaching to devise activities that will support the development of phonological concepts.

For example, having observed my learners' difficulties with syllable codas such that *drunk* and *drunker* sounded the same to them I developed an information gap activity, which would help to make the difference in meaning and form clear and give them the opportunity to practise and receive feedback multiple times. I called this the "Drunk Snail Game" and it involves sets of cards containing pairs of adjectives and their comparatives, describing animals such as: a drunk snail/a drunker snail; a loud parrot/a louder parrot, etc. I printed each item on a separate card with an appropriate picture taken from clip art. The object of the game is to find matching pairs by correctly pronouncing what is on the card. Another player who has the matching card then has to correctly pronounce what is on that card. The players check that they have understood each other by comparing their cards. If they are the same, they have succeeded and, if not, one of them will realize that they pronounced it incorrectly. The details of this game are described in Couper (2014). This is an example of how we can develop activities that focus on concept formation by establishing appropriate figure-ground organization, helping learners to establish category boundaries through the cognitive ability to compare and detect discrepancy and learn from feedback, and providing multiple experiences in a context that presents learning as social behavior.

Most other communicative activities can also be structured with the aim of helping learners to form and practise new concepts. For example, in setting up a role-play activity we might first consider the type of language that will be needed and possibly have some controlled practice with the language beforehand, that is, the lexico-grammatical aspects as well as the phonological ones. One could record the role-play and review it with the class, allowing them to focus on their performance and discussing certain pronunciation features that caused misunderstandings. This might then lead on to further practice once learners have understood the form-meaning relationship. Another common task for learners is

to give oral presentations. Again, we can prepare for this by focusing on certain key features such as phrasing and sentence stress. Then after their performance, one could review the video leading on to increased awareness as well as plans for further practice. These sorts of activities are already common in the communicative language classroom. We can easily integrate a pronunciation focus by thinking about what helps learners to practise and form new concepts.

Teaching tip five: provide corrective feedback focused on concept formation

Corrective feedback (CF) is the most common way in which teachers engage with pronunciation (Foote et al. 2013); therefore it is important to consider how it is provided. A key factor in determining the effectiveness of CF is the extent to which the learner understands the correction. The first step then is to make sure learners understand that it is a correction and that they understand precisely where the problem lies. As an example of how easy it is for corrective feedback to miss the mark, Couper (2013: 10) reports the following event:

> In the practice stage, during which key words from the listening were being practised, Ay repeated "exports" for "experts". The teacher then repeated "experts" several times while Ay continued to say "exports". This was being done without any visual support on the board. Bea then explained to Ay that the /p/ changes to /b/ and Ay commented that "when I say my name everyone thinks it's a 'p' but it's a 'b'". The teacher finally realized that Ay had been focusing on the wrong thing (i.e. the pronunciation of /p/ rather than the second vowel in "experts") and wrote the two words on the board, underlining the vowel. Ay then saw where the problem was and attempted to correct it, although she still found it difficult.

This reminds us of the need to write things on the board to help learners see where their mistakes are. Of course, even when they know where the problem is it does not always mean they can fix it. Cognitive Linguistics makes it clear that if learners do not understand the phonological concept they are not likely to learn from the correction. So we have to ensure that we communicate effectively with our students when making corrections. This is where SCM comes in. If we have already developed this in relation to the particular phonological concept that is the focus of correction, then we should be able to quickly communicate what the problem is. For example, when correcting syllable codas, rather than referring to the additional syllable, or even the additional vowel, we can use the learners' descriptions such as "that's too strong", "say it shorter", "it becomes quiet", "make it smaller" (Couper 2013: 9). If they are still unable to correct it then they need further practice, possibly using Critical Listening techniques, to fully establish the concept. In conclusion, the teacher must focus on providing CF that contributes to the formation of phonological concepts.

Lyster, Saito, and Sato (2013) in their review of research into oral CF note that there is general support for CF from a range of theoretical perspectives; however

meta-analyses of research into the effectiveness of CF tend to categorize it according to whether it is explicit or implicit, a prompt or a reformulation, or whether it is form focused or meaning focused. The position put forward in this paper is that these variables fail to isolate the most important variable, which is the degree to which the communication between the teacher and student is actually understood.

Conclusion

To sum up, this approach suggests we need to define instruction in terms of how it may help concept formation. Students can achieve this by accessing a range of cognitive abilities and applying them in a social setting that permits the co-construction of meaning. This leads to an understanding of the link between form and meaning. The teacher can foster this process by raising awareness of differences in perception between the L1 and L2. Most importantly, this involves explicit and meaningful communication about these differences through the social construction of metalanguage. As teachers we also need to help form category boundaries by providing opportunities for students to compare and contrast a production that will be perceived correctly with one that will not. This involves Critical Listening. Students need to be actively involved in the meaning-making process such as would be expected in a broadly communicative approach to language teaching. Thus classroom activities only need to be adjusted slightly to allow for SCM and Critical Listening to ensure that they lead to effective concept formation. Once SCM has been established, effective corrective feedback will follow much more easily. As has been seen, this approach is not in disagreement with positions from a number of theoretical positions, especially those that espouse a participant metaphor. For example, Trofimovich, Kennedy, and Foote (Chapter 20 in this issue) note that pronunciation learning is a "complex sociocognitive and situationally embedded phenomenon", a view very much in line with the tenets of Cognitive Linguistics.

REFERENCES

Acton, W., Baker, A.A., Burri, M., and Teaman, B. 2013. Preliminaries to haptic-integrated pronunciation instruction. In: *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference*, J. Levis and K. LeVelle (eds.), 234–244, Ames, IA: Iowa State University.

Anderson, J. 1993. *Rules of the Mind*, Hillsdale, NJ: Erlbaum.

Anderson, J. 2000. *Cognitive Psychology and Its Implications*, New York: Worth Publishers.

Atkinson, D. 2011. A sociocognitive approach to second language acquisition: how mind, body, and world work together in learning additional languages. In: *Alternative Approaches to Second Language Acquisition*, D. Atkinson (ed.), 143–166, New York: Routledge.

Block, D. 2003. *The Social Turn in Second Language Acquisition*, Edinburgh: Edinburgh University Press.

Blyth, C. 1997. A constructivist approach to grammar: teaching teachers to teach aspect. *The Modern Language Journal* 81**:** 50–66.

Bybee, J. 2001. *Phonology and Language Use*, Cambridge: Cambridge University Press.

Carr, J. 1999. From "sympathetic" to "dialogic" imagination: cultural study in the foreign language classroom. In: *Striving for the Third Place: Intercultural Competence through Language Education*, J. Lo Bianco, A. Liddicoat, and C. Crozet (eds.), Melbourne: Language Australia.

Chomsky, N. and Halle, M. 1968. *The Sound Pattern of English,* New York: Evanston, and London: Harper & Row.

Couper, G. 2006. The short and long-term effects of pronunciation instruction. *Prospect* 21**:** 46–66.

Couper, G. 2011. What makes pronunciation teaching work? Testing for the effect of two variables: socially constructed metalanguage and critical listening. *Language Awareness* 20(3): 159–182. doi: 10.1080/09658416.2011.570347.

Couper, G. 2012. Teaching word stress: learning from learners' perceptions. *TESOL in Context* S3 (November 2012). Available at: http://www.tesol.org.au/Publications/Special-Editions.

Couper, G. 2013. Talking about pronunciation: socially constructing metalanguage. *English Australia* 29(1): 3–18.

Couper, G. 2014. Teaching concepts of pronunciation: syllables, stress and drunk snails. *Speak Out!* 50: 34–38.

Crozet, C. and Liddicoat, A. 1999. The challenge of intercultural language teaching: engaging with culture in the classroom. In: *Striving for the Third Place: Intercultural Competence through Language Education*, J. Lo Bianco, A. Liddicoat, and C. Crozet (eds.), Melbourne: Language Australia.

Cruttenden, A. 2001. *Gimson's Pronunciation of English*, London: Arnold.

DeKeyser, R. 1998. Beyond focus on form: cognitive perspectives on learning and practicing second language grammar. In: *Focus on Form in Classroom Second Language*, C. Doughty and J. Williams (eds.), Cambridge: Cambridge University Press.

Doughty, C. 2003. Instructed SLA: constraints, compensation, and enhancement. In: *The Handbook of Second Language Acquisition*, C. Doughty and M. Long (eds.), Blackwell.

Ellis, N. 2001. Memory for language. In: *Cognition and Second Language Instruction*, P. Robinson (ed.), Cambridge: Cambridge University Press.

Ellis, R. 2002. Direct intervention in language learning. *New Zealand Studies in Applied Linguistics* 8**:** 45–61.

Ellis, R. 2006. Current issues in the teaching of grammar: an SLA perspective. *TESOL Quarterly* 40**:** 83–107.

Ellis, R. 2009. Implicit and explicit learning, knowledge and instruction. In: *Implicit and Explicit Knowledge in Second Language Learning, Testing and Teaching*, R. Ellis and S. Loewen (eds.), 3–25, Clevedon: Channel View Publications.

Flege, J. 1990. English vowel production by Dutch talkers: more evidence for the "similar" vs. "new" distinction. In: *New Sounds 90 Conference*, J. Leather and A. James (eds.), Amsterdam: University of Amsterdam.

Flege, J. 1995 Second language speech learning: theory, findings, and problems. In: *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, W. Strange (ed.), Baltimore, MD: York Press.

Foote, J.A., Trofimovich, P., Collins, L., and Urzúa, F.S. 2013. Pronunciation teaching practices in communicative second language classes. *The Language Learning Journal*. doi: 10.1080/09571736.2013.784345.

Fraser, H. 2004. *Teaching Pronunciation: A Guide for Teachers of English as a Second Language* (CD-ROM), Canberra: DETYA.

Fraser, H. 2006. Phonological concepts and concept formation. *IJES, International Journal of English Studies* 6(2): 55–76.

Fraser, H. 2009. Pronunciation as categorization: the role of contrast in teaching English /r/ and /l/. In: *Studies in Applied Linguistics and Language Learning*, A. Mahboob and C. Lipovsky (eds.), 289–306, Newcastle upon Tyne: Cambridge Scholars Publishing.

Fraser, H. 2010. Cognitive phonology as a tool for teaching pronunciation. In: *Fostering Language Teaching Efficiency Through Cognitive Linguistics*, S. de Knop, F. Boers, and T. De Rycker (eds.), 357–380, Berlin: Mouton de Gruyter.

Fromkin, V., Rodman, R, Collins, P., and Blair, D. 1996. *An Introduction to Language*, Sydney: Harcourt Brace.

Gibbons, P. 2006. *Bridging Discourses in the ESL Classroom*, London and New York: Continuum.

Gregg, K. 2001. Learnability and second language acquisition theory. In: *Cognition and Second Language Instruction*, P. Robinson (ed.), Cambridge: Cambridge University Press.

Guion, S. and Pederson, E. 2007. Investigating the role of attention in phonetic learning. In: *Language Experience in Second Language Speech Learning*, O. Bohn and M. Munro (eds.), Amsterdam and Philadelphia: John Benjamins.

Holme, R. 2012. Cognitive linguistics and the second language classroom. *TESOL Quarterly* 46(1): 6–29.

Housen, A. and Pierrard, M. 2005. Investigating instructed second language acquisition. In: *Investigations in Instructed Second Language Acquisition*, A. Housen and M. Pierrard (eds.), Berlin: Mouton de Gruyter.

Jordan, G. 2004. *Theory Construction in Second Language Acquisition*, Amsterdam and Philadelphia: John Benjamins.

Krashen, S. 1982. *Principles and Practice in Second Language Acquisition*, Oxford: Pergamon Press.

Kuhl, P. and Iverson, P. 1995. Linguistic experience and the "perceptual magnet effect". In: *Speech Perception and Linguistic Experience: Issues in Cross Language Research*, W. Strange (ed.), Timonium, MD: York Press.

Langacker, R. 1987. *Foundations of Cognitive Grammar*, vol. 1, *Theoretical Prerequisites*, Stanford: Stanford University Press.

Lantolf, J. 1996. SLA theory building: "Letting all the flowers bloom". *Language Learning* 46**:** 713–749.

Lantolf, J. 2011. Integrating scoiocultural theory and cognitive linguistics in the second language classroom. In: *Handbook of Research in Second Language Teaching and Learning:* vol. 2, E. Hinkel (ed.), 303–318, Hoboken, NJ: Routledge.

Lantolf, J. and Thorne, S. 2006. Introduction. In: *Sociocultural Theory and the Genesis of Second Language Development*, J. Lantolf and S. Thorne (eds.), Oxford: Oxford University Press.

Leather, J. 1999. Second language speech research: an introduction. In: *Phonological Issues in Language Learning*, J. Leather (ed.), Oxford: Blackwell.

Linell, P. 2005. *The Written Language Bias in Linguistics*, London and New York: Routledge.

Lively, S., Pisoni, D., and Goldinger, S. 1994. Spoken word recognition: research and theory. In: *Handbook of Psycholinguistics*, Academic Press.

Lo Bianco, J., Liddicoat, A., and Crozet, C. 1999. *Striving for the Third Place: Intercultural Competence through Language Education*, Melbourne: Language Australia.

Long, M. 1996. The role of the linguistic environment in second language acquisition. In: *Handbook of Second Language Acquisition*, W. Ritchie and T. Bhatia (eds.), San Diego, CA: Academic Press.

Lyster, R., Saito, K., and Sato, M. 2013. Oral corrective feedback in second language classrooms. *Language Teaching* 46: 1–40. doi: 10.1017/S0261444812000365.

Macaro, E. 2003. *Teaching and Learning a Second Language*, London and New York: Continuum.

Mompean, J.A. 2006. Introduction: cognitive phonology in cognitive linguistics. *International Journal of English Studies* 6: vii–xii.

Mompean, J.A. 2014. Phonology. In: *Bloomsbury Companion to Cognitive Linguistics*, J.R. Taylor and J. Littlemore (eds.), 357–392, Bloomsbury.

Munro, M. and Bohn, O. 2007. The study of second language speech. In: *Language Experience in Second Language Speech Learning*, O. Bohn and M. Munro (eds.), Amsterdam and Philadelphia: John Benjamins.

Murphy, G.L. 2002. *The Big Book of Concepts*, Cambridge, MA: MIT (Bradford).

Norris, J. and Ortega, L. 2001. Does type of instruction make a difference? Substantive findings from a meta-analytic review. *Language Learning* 51: 157–213.

Ortega, L. 2011. SLA after the social turn: where cognitivism and its alternatives stand. In: *Alternative Approaches to Second Language Acquisition*, D. Atkinson (ed.), 167–180, New York: Routledge.

Pisoni, D. and Lively, S. 1995. Variability and invariance in speech perception: a new look at some old problems in perceptual learning. In: *Speech Perception and Linguistic Experience: Issues in Cross Language Research*, W. Strange (ed.), Timonium, MD: York Press.

Rochet, B. 1995. Perception and production of second-language speech sounds by adults. In: *Speech Perception and Linguistic Experience: Issues in Cross Language Research*, W. Strange (ed.), Timonium, MD: York Press.

Sanz, C. 2005. Adult SLA: the interaction between external and internal factors. In: *Mind and Context in Adult Second Language Acquisition*, C. Sanz (ed.), Washington, DC: Georgetown University Press.

Schmidt, R. 2001. Attention. In: *Cognition and Second Language Instruction*, P. Robinson (ed.), Cambridge: Cambridge University Press.

Shockey , L. 2003. *Sound Patterns of Spoken English*, Oxford: Blackwell.

Strange, W. (ed.). 1995. *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, York Press.

Swain, M. 1995. Three functions of output in second language learning. In: *Principle and Practice in Applied Linguistics: Studies in Honour of H.G. Widdowson*, G. Cook and B. Seidlhoffer (eds.), Oxford: Oxford University Press.

Swain, M. 2000. The output hypothesis and beyond: mediating acquisition through collaborative dialogue. In: *Sociocultural Theory and Second Language Learning*, J. Lantolf (ed.), Oxford, Oxford University Press.

Swain, M. 2006. Languaging, agency and collaboration on advanced language proficiency. In: *Advanced Language Learning: The Contribution of Halliday and Vygotsky*, H. Byrnes (ed.), 95–108, London: Continuum.

Taylor, J. 2002. *Cognitive Grammar*, Oxford: Oxford University Press.

Thorne, S. 2000. Second language acquisition theory and the truth(s) about relativity. In: Lantolf, J. (Ed.) *Sociocultural Theory and Second Language Learning*, J. Lantolf (ed.), Oxford: Oxford University Press.

VanPatten, B. 2010. Some verbs are more perfect than others: why learners have difficulty with ser and estar and what it means for instruction. *Hispania* 93(1): 29–38.

Warren, R. 1982. *Auditory Perception: A New Synthesis*, New York: Pergamon Press.

Widdowson, H. 2003. *Defining Issues in English Language Teaching*, Oxford: Oxford University Press.

Williams, M. and Burden, R. 1997. *Psychology for Language Teachers: A Social Constructivist Approach*, Cambridge: Cambridge University Press.

Zuengler, J. and Miller, E. 2006. Cognitive and socio-cultural perspectives: two parallel SLA worlds? *TESOL Quarterly* 40**:** 35–58.

24 The Pronunciation of English as a Lingua Franca

ROBIN WALKER AND WAFA ZOGHBOR

Introduction

For most English language teachers the current goal of pronunciation teaching is either (near-) native-speaker competence or "comfortable intelligibility" (Kenworthy 1987). Both goals assume a native-speaker listener and nobody seriously questioned the dependence of pronunciation teaching on NS norms until the publication of empirical data obtained from the study of interactions between non-native speakers (Jenkins 1998). These data led Jenkins to challenge the validity of native-speaker judgments of intelligibility when English was being used as a lingua franca (ELF). Her research also led to the development of the *Lingua Franca Core* (LFC), a set of key segmental and suprasegmental pronunciation features present in all NNS English regardless of the speaker's accent (Jenkins 2000).

This chapter will explore the origins of the pronunciation for English as a lingua franca, before going on to detail how an ELF approach to teaching pronunciation can be put into practice. Key issues that will be dealt within the chapter include:

- ENL, ESL, EFL, and ELF: differences in pronunciation teaching goals.
- Variation, accent, and intelligibility.
- The Lingua Franca Core.
- Teaching ELF pronunciation – classroom models.
- Teaching ELF pronunciation – classroom techniques.
- Teaching ELF pronunciation – the learner's mother-tongue phonology.
- Concerns regarding the teaching of pronunciation for ELF.

The Handbook of English Pronunciation, First Edition. Edited by Marnie Reed and John M. Levis.

ENL, ESL, EFL and ELF: differences in pronunciation teaching goals

Kachru (1985) conceptualized the spread of English in three circles. The inner circle represents the countries where English is used as a native language (ENL) by those who have traditionally been described as its native speakers (the United States, the United Kingdom, Canada, Ireland, Australia, and New Zealand). The outer circle represents countries where English has a colonial history and is now used as a second language (ESL) alongside other official languages (Malaysia, Singapore, and Kenya, for example). The expanding circle represents the remaining countries where English is used as a foreign language (EFL), that is to say, in order to facilitate communication with the language's native speakers in the inner circle countries.

English as a lingua franca (ELF) is not another circle to be added to Kachru's model. Rather it refers to the ways in which English is now being used within the three circles. Most frequently, English is being used as a lingua franca between members of two or more expanding circle countries who do not share the same first language (Jenkins 2007; Seidlhofer 2005). ELF interactions are not normally defined to include, although they do not exclude, native speakers. Moreover, ELF interactions can occur within the inner circle itself, as is the case when communication is between non-native speakers visiting, working or studying in an inner circle country.

Table 24.1 summarizes the four settings. The first three (ENL, ESL, and EFL) correspond to the use of English in Kachru's three circles. The fourth shows how ELF differs from the other three.

These different settings have implications on teaching/learning pronunciation in three dimensions:

- the specific phonological features to be included in a pronunciation syllabus.
- the inclusion of accommodation skills as an essential requirement in communication among interlocutors (Jenkins 2000).
- the way in which learners' errors are perceived.

Variation, accent, and intelligibility

Speakers of the same language vary in the way they speak. This variation might be due to geographical distance, social variables, or through the language evolving to meet the needs of its speakers. If variation is found across the entire linguistic system – grammar, vocabulary, pragmatics, and pronunciation – we refer to it as a dialect. Accent, in contrast, refers to variation exclusively in pronunciation. It is perfectly feasible for two speakers to use the same dialect of English with different accents. Standard English, for example, is widely spoken throughout the United Kingdom by speakers from upper-class and upper-middle-class backgrounds.

Table 24.1 Differences in using English in ENL, ESL, EFL, and ELF contexts.

	ENL	*ESL*	*EFL*	*ELF*
Users	Traditionally referred to as the native speakers	Speakers of English as a second language	Speakers of English as a foreign language	Users of English as a lingua franca
Location	Kachru's inner circle	Kachru's outer circle	Kachru's expanding circle	No specific location; mainly involves speakers from the expanding and outer circles
Norms	Norm providing: has long been the model for speakers from the expanding circle	Norm developing: has developed its own varieties; does not depend on NS norms and models	Norm-dependent: correctness in pronunciation based on NS norms	Uses most features of NS Englishes, but accepts all intelligible variations
Function	Communication between native speakers	Communication in English within an ESL context	Communication with the NSs of English	International communication, predominantly between NNSs

However, as Trudgill points out, "[p]erhaps 9%–12% of the population of Britain … speak Standard English with some form of regional accent" (Trudgill 1999: 118).

Accents, then, are a natural and inevitable outcome of language variation. They are also one of the most salient aspects of variation, to the point that they are frequently used to classify speakers, either geographically, socially, or both. In this respect, it is not uncommon to hear language teachers and lay people refer to learners as having a foreign accent when referring to differences in NNS pronunciation compared to an NS norm. However, such a reference is problematic when English is being used as a lingua franca, since ELF, by definition, has no foreigners. Thus, whilst the goal of a great deal of pronunciation teaching in EFL, explicit or otherwise, is the (near) elimination of the learner's foreign accent, the goal of pronunciation for learners of English as a lingua franca is mutual intelligibility across an ever-widening range of accents.

Although intelligibility has long been acknowledged as a key issue in pronunciation, linguists do not subscribe to a single definition. Smith and Nelson (1985) referred to intelligibility as the ability of the listener to recognize individual words or utterances, whilst defining comprehensibility as the listener's ability to understand the meaning of the word or utterance in its given context. In contrast, Derwing and Munro (Munro and Derwing 1995; Derwing and Munro 1997) define intelligibility as the extent to which a speaker's utterance is actually understood, whilst for them comprehensibility refers to the listener's estimation of the difficulty or ease of understanding an utterance.

The work of Derwing and Munro, like that of Smith and Nelson, stresses the importance of the distinction between intelligibility and comprehensibility, though, for both groups, being able to do well in one of the two areas does not ensure success in the other (Derwing 2006; Smith and Nelson 1985). Nelson points out that "comprehensibility can fail even when the degree of intelligibility between participants is high" (Nelson 2008: 302). Zielinski (2004), for example, found that listeners could identify individual words accurately but puzzle over the whole message. Matsuura, Chiba, and Fujieda (1999), on the other hand, found that although Japanese listeners could easily understand the utterances in their study, they could not transcribe the words correctly, transcription being a standard test of intelligibility.

In their work on intelligibility, Derwing and Munro (1997) referred to the term "accentedness", which they use in order to indicate the degree to which a particular accent differs from a local norm. Accentedness, they concluded, is quite different from intelligibility: "One very robust finding in our work is that accent and intelligibility are not the same thing. A speaker can have a very strong accent, yet be perfectly understood" (Derwing and Munro 2008: 1). The distinction they make between accent and intelligibility is crucial to ELF pronunciation, given the goal of mutual intelligibility across the range of accents that characterize ELF encounters.

A key assumption of the research on intelligibility is the belief that it is not a one-way process, that the burden to make him- or herself intelligible to the listener(s) does not lie exclusively with the speaker. For Smith and Nelson, for example, "intelligibility is not speaker- or listener-centered, but is interactional between speaker and listener" (Smith and Nelson 1985: 333). For too long, listening

has been described as a receptive skill, when in practice speaker–listeners make themselves intelligible to each other by co-constructing meaning. Jenkins (2000) places great importance on this process of negotiation of meaning, and in particular highlights the role of accommodation as a central skill for successful ELF exchanges.

As with intelligibility, accommodation is not a simple concept. *Communication Accommodation Theory* (CAT) (Giles and Coupland 1991) holds that people's verbal and nonverbal behavior can change according to the setting, the topic, or the interlocutor. CAT interprets the way people attune to others during an interaction by using three strategies: *convergence*, whereby individuals shift towards each other's communicative behaviors; *divergence*, which refers to how speakers accentuate speech and nonverbal differences between themselves and others; and *maintenance*, whereby interactants preserve their speech patterns and other communicative behaviors in order to maintain their group identity (Giles, Coupland, and Coupland 1991). Jenkins (2000) offers an excellent introduction to accommodation theory, and in particular describes how phonological accommodation may be motivated by:

1. Solidarity amongst speakers, leading the pronunciation patterns of interlocutors to convergeing on each other.
2. Communicative efficiency – adjustments in pronunciation made to facilitate communication, and also involving speech patterns converging.
3. Identity maintenance – the preservation of speech patterns by a speaker in order to reinforce membership of a group external to the communication in hand, possibly leading to diverging speech patterns among interlocutors.

Motivation 3 does not require speakers to make any changes in their pronunciation, and so is of no interest here. Motivation 1 is interesting, but is only likely to come about when speakers work or live together on a long-term basis. In practice, however, most ELF discourse occurs through short-term interactions between interlocutors who are not yet fully competent. In such situations Motivation 2, communicative efficiency, is the driving force behind any attempts at accommodation.

In ELF interactions, changes that are deliberately made to a speaker's pronunciation constitute an important accommodation strategy, and Jenkins' data (2000) offer s clear examples of speakers modifying their pronunciation in order to make themselves more intelligible to their interlocutors. In particular, they employed phonological accommodation in order to converge on a common ground of mutually intelligible English, and it is to what constitutes this common phonological ground that we now turn our attention.

The Lingua Franca Core

Pronunciation targets adopted in English language teaching are generally derived from native-speaker varieties of English, principally the standard British and American English varieties of Received Pronunciation (RP) and General American

(GA) respectively. In an attempt to provide similar targets for ELF pronunciation, Jenkins identified empirically which phonological features caused breakdowns in NNS–NNS communication, and revised the contents of existing pronunciation syllabuses to generate the Lingua Franca Core (LFC), the segmental and supra-segmental features required for intelligible spoken communication among NNSs (Jenkins 2000). The inclusion or exclusion of certain features from the LFC was based essentially on two criteria: their influence on intelligibility among NNS interlocutors and the concept of teachability–learnability. For Dalton and Seidlhofer (1994) "[s]ome things, say the distinction between fortis and lenis consonants, are fairly easy to describe and generalize – they are teachable" (1994: 72–73). In contrast, other areas of English pronunciation, because of their complex nature, are only learnable, i.e., acquirable, outside the classroom. For example, "the attitudinal use of intonation is something that is best acquired through talking with and listening to English speakers" (Roach 1991: 169).

Table 24.2 compares EFL pronunciation targets with the LFC. Column B lists the generally agreed EFL pronunciation targets, whilst column C indicates the impact of these features in ELF communication. Column D shows the targets for ELF pronunciation.

The core features of the LFC are:

1. The consonant inventory
   All the consonant sounds of the RP/GA syllabus are core features of EFL pronunciation. One very significant exception to this are the dental fricatives /θ/ and /ð/. Absent from many of the world's languages, as well as from a number of NS varieties and regional accents, these two consonants are especially resistant to classroom teaching (Jenkins 2000; Pennington 1996). Moreover, certain substitutions of /θ/ and /ð/ are found to be fully intelligible in ELF contexts. The commonest of these are the dental plosives [t̪] and [d̪] and the labio-dental fricatives /f/ and /v/. A third substitution is that of /s/ for /θ/ and /z/ for /ð/. A preference, rather than an exception, is that of the RP intervocalic [t] over the GA intervocalic flap /ɾ/ when <t> occurs intervocalically, as in words like "water" or "matter". This is because of the proximity of the GA variant to /d/, which can result in "matter", for example, sounding like "ma*dd*er".
2. Phonetic requirements
   Aspiration. In the absence of aspiration following the fortis plosives /p/, /t/, and /k/ in the initial position in a stressed syllable, the listener will find it more difficult to identify /p/, /t/, and /k/ as voiceless. An unaspirated /p/ may be mistaken for /b/, unaspirated /t/ for /d/, and /k/ for /g/, with "peach" sounding like "beach", and so on (Jenkins 2000; Osimk 2009; Rajadurai 2006; Zoghbor 2011b).
   Vowel length. There is a marked shortening of vowel length before fortis consonants (fortis clipping). The long vowel /iː/ is shorter in "seat" than in "seed", for example, and the /æ/ is shorter in "back" than in "bag". This phonetic effect is seldom included in EFL pronunciation syllabuses, perhaps because it is considered an aspect of pronunciation for advanced-level learners only, but is a core feature for ELF.

Table 24.2 Pronunciation targets for teaching EFL and ELF (from Zoghbor 2011a, modified from Jenkins 2005).

#	*A*	*B*	*C*	*D*
	Aspect of pronunciation	*EFL targets*	*Impact on ELF intelligibility*	*ELF targets*
1	The consonant inventory	All sounds	√ but not all	All sounds except /θ/ and /ð/ and the allophone [ɬ]
		RP nonrhotic /r/ GA rhotic /r/	Not clear	Rhotic /r/ preferred
		RP intervocalic [t] GA intervocalic [ɾ]	Not clear	Intervocalic [t] preferred
2	Phonetic requirements	Rarely specified	√ but not all	Aspiration after /p/, /t/, and /k/. Appropriate vowel length before fortis/lenis consonants.
3	Consonant clusters	All word positions	√ but not all	Word initially, word medially
4	Vowel quantity	Long–short contrast	√	Long–short contrast
5	Vowel quality	Close to RP or GA	X	Consistent L2 regional qualities acceptable
6	Weak forms	Important	X	Unhelpful to intelligibility
7	Other features of connected speech	Important	X	Inconsequential or unhelpful
8	Stress-timed rhythm	Important	X	Inconsequential
9	Word stress	Important	X	Unnecessary/can reduce flexibility
10	Nuclear (tonic) stress	Important	√	Important
11	Pitch movement	Important	X	Inconsequential

3. Consonant clusters
   Consonant clusters in the word-initial position are a core feature. Speakers from L1s with relatively few clusters, or with an underlying consonant–vowel syllable structure, simplify clusters, either by the addition of a vowel or by deletion of one or more of the consonants. Of these two strategies, addition is less damaging to intelligibility in ELF than deletion. Thus, while the pronunciation of "sting" as [eˈstɪŋ] or "stone" as [sɪˈtəʊn] is not found to threaten ELF intelligibility, the deletion of one of the consonants does, since it produces either [sɪŋ] or the nonsense word [tɪŋ], which might be understood as "thing".
4. Vowel sounds
   The maintenance of the contrast between long and short vowels, such as those found in "feel" and "fill" or "pool" or "pull", is a core feature. Vowel quality, in contrast, is non-core. In ELF contexts, variations in vowel quality that are *consistent* are found to be intelligible, and are seen as a regional variation on a par with that which exists in the vowel qualities of different NS Englishes. Jenkins (2000) highlights one exception. In her data, the quality as well as the quantity of the long central vowel /ɜː/ was found to be important for ELF intelligibility.
5. Nuclear stress
   The nuclear stress carries the most salient part of the speaker's message, and thus the focus of the listener's attention. Deviations in the placement of the nucleus have the potential to affect the listener's processing of entire chunks of the message. Jenkins (2000) gives the example of a Swiss speaker explaining to her Taiwanese interlocutor how many cigarettes she smoked a day. The Taiwanese speaker responded:

   you smoke more than i DO

   The speaker was comparing her smoking habits with those of her interlocutor, and her failure to place the nuclear stress on "I" to signal contrast meant that she had to repeat the utterance several times before her interlocutor was able to understand it. Jenkins argues that the rules of unmarked and contrastive are simple enough for learners to master, can easily be integrated receptively and productively into classroom work, and operate at a more conscious level than the other aspects of the intonation system such as pitch movement.

A number of features have been described as having no influence on intelligibility for ELF speakers (Deterding 2010; Jenkins 2000; Zoghbor 2011b).

1. Dark /l/
   - The velarized lateral approximant [ɫ] commonly known as dark "l", is regularly substituted with either /l/ (clear "l") or /ʊ/ in the speech of both NNSs and NSs, leading to "milk" being heard as [mɪɫk], [mɪlk], or [mɪʊk]. The majority of RP speakers, for example, pronounce the pre-consonant dark /l/ as /ʊ/ in noncareful speech. Neither substitution is problematic

for ELF intelligibility. In contrast, dark /l/ is problematic for most learners in production. Many never acquire it.

2. Post-alveolar approximant [ɹ]
   - The LFC opts for rhotic variants like those in GA rather than nonrhotic variants like RP, "mainly because the 'r' is indicated orthographically in situations (i.e. post-vocalic) where it does not feature in the RP pronunciation when a word is spoken in isolation (for example, 'four') or is followed by another consonant …" (Jenkins 2000: 139). Jenkins has argued that adopting a rhotic variant that better reflects the spelling should increase ELF intelligibility. Research by Osimk (2009) neither supports this nor refutes it. With respect to which variant of /r/ is best understood, Zoghbor (2011b) found that the trill used by her Arabic-L1 subjects caused no problems of intelligibility. For Walker (2010), the most likely variant of /r/ to be heard in ELF is the trill, whilst the only variant that might not be widely intelligible is the uvular [ʁ] characteristic of French and southern German accents of English.
3. Word stress
   - Word stress is considered non-core in the LFC, but at the same time "is something of a grey area" (Jenkins 2000: 150). On the one hand, the rules for predicting word stress are so complex that they are argued to be almost not acquirable, an idea that has also been documented by Brown (1992) and used by such authors as Nasr (1963), Zawaydeh *et al*. (2002), Kharma and Hajjaj (1997), and Benrabah (1997) to account for the difficulty of learning English word stress by L2 learners. On the other hand, misplaced word stress has a corresponding effect on nuclear stress and, as we have already seen, misplaced nuclear stress has a serious effect on ELF intelligibility.
4. Stress timing
   - The distinction between stress timing and syllable timing is not clear, and may be more pedagogic than real (Marks 1999; Roach 1991). More importantly, there is no empirical data to suggest that the lack of stress timing affects intelligibility in ELF. As a result, "there seems little need for learners of English around the world to adopt this approach, given that syllable-based rhythm is so widespread in varieties of World English and in many cases it seems to enhance intelligibility" (Deterding 2010: 10).
5. Pitch movement
   - Some experts now feel that the multiple and complex ways in which NSs use pitch movement (tone) are simply not teachable in the language classroom. Roach points out that "the complexity of the total set of sequential and prosodic components of intonation and of paralinguistic features make it a very difficult thing to teach or learn" (Roach 2009: 151). Cauldwell goes even further when he declares that "after working for nearly twenty years with Discourse Intonation on examples of spontaneous speech I no longer feel that tones 'mean' anything" (Cauldwell 2006). In addition, there are no data to suggest that poor selection of tone impacts negatively on intelligibility in ELF.

Certain features of NS pronunciation, particularly weak forms, schwa, and vowel reduction, are not only unnecessary for intelligibility in ELF settings; they can actually have a negative impact. Weak forms, for example, hold potential problems of "recoverability" where NNSs interlocutors are unable "to work backward from the surface form through a derivation to obtain the unique underlying representation" (Weinberger 1987: 404). The arguments against teaching schwa and vowel reduction are similar. Avery and Ehrlich (1992) pointed out that the relative absence of reduced vowel sounds did not seem to cause any misunderstanding, whilst Deterding (2010) argued that "avoiding reduced vowels is the norm in new varieties of English around the world, and speakers of such Englishes find that the use of full vowels in function words can enhance intelligibility" (2010: 9).

Teaching ELF pronunciation – classroom models

Although the goals and prioritized features of an ELF approach to pronunciation are clear, as we have seen in the previous two sections of the chapter, a key issue in an ELF-based approach to pronunciation teaching is the choice of a model. In EFL, with an NS accent as the goal, the model is a speaker of that accent. Choosing a model for an ELF approach is more complex; NS standard accents are not directly relevant to ELF goals and ELF is spoken with a vast range of accents. Thus, for ELF, the term "model" encompasses any speaker, with any accent, who at a minimum is competent in the features of the LFC. This speaker can be a native speaker of English, but given the demographics of ELF, is most likely to be a non-native speaker.

At a purely pragmatic level, three options are available to teachers when providing a classroom model within an ELF approach:

1. **Existing native-speaker materials.** At the time of writing, the ELT market is characterized by the almost complete absence of pronunciation materials that employ ELF voices as models. Thus, while teachers wait for ELF-specific materials to become available, they are almost obliged to use those based on one or another standard NS accent. The only precaution if they do this is to minimize working on those non-core features that have been identified as either not helpful or as potentially damaging to ELF intelligibility.
2. **Competent ELF users.** The alternative to employing selected features of an NS accent is to use the accent of a competent ELF user. In an appraisal of suitable models for teaching ELF, Ur argues that "the model for ELF teachers should be the *fully competent ELF user*, without defining whether such a speaker was or was not originally a "native speaker" (Ur 2010: 87 – italics in original). Seidlhofer (2011) also strongly supports this approach and goes on to suggest that it has a significant advantage for the learner. Unfortunately, although (fully) competent users of ELF abound in international politics, academia, entertainment, and sport, recordings of such speakers are still noticeably absent from current published ELT materials.

3. **The teacher.** Teachers have always been models for learners. This is also true for ELF, and teachers who know from personal experience that their pronunciation is intelligible in ELF contexts can confidently act as models for their learners. Indeed, if we look back over the history of ELT, we will see that in practice countless NS teachers have successfully acted as models even though they do not speak English with a standard accent. ELF merely extends this prerogative to NNS teachers; in ELF contexts accents are valid if they are intelligible, rather than because of their origin and status.

In practice, until ELT materials are available with competent ELF users as models, teachers will struggle with Option 2, although Walker (2010) indicates a number of ways to alleviate the problem. This leaves teachers with Options 1 and 3. Of these, although a less confident teacher will probably be more comfortable with the former, there are good reasons for promoting the third option. Clarifying a common misinterpretation of an ELF approach to pronunciation, Jenkins insists that the model "is not the LFC but the local teacher whose accent incorporates both the core features and the local version of the non-core features" (Jenkins 2007: 25). One significant outcome of Option 3 is the empowerment it supposes of the local (bilingual) NNS teacher, who is placed on (at least) equal footing with NS teachers of English.

Teaching ELF pronunciation – classroom techniques

An ELF approach to teaching pronunciation centers around two areas – competence in the LFC and good accommodation skills – and two major teaching situations – multilingual groups, with students from a range of different L1 backgrounds, and monolingual groups, where the L1 background is shared by the students, and usually the teacher.

1. Competence in the LFC
   A great deal of what can be found in existing pronunciation manuals can be applied directly to the teaching of pronunciation for ELF, although certain features of English pronunciation that are central to ELF phonological competence are often only considered suitable for advanced learners of English as a foreign language. This is the case with the aspiration of the voiceless (fortis) plosives /p, t, k/ or with fortis-clipping (the shortening of vowels when followed by voiceless consonants). It is also largely true for the treatment of word-initial consonant clusters or nuclear stress placement, both of which would need to be tackled early in an ELF approach.

   Another feature of teaching pronunciation for ELF is the extent to which the learner's L1 phonology can be brought to bear on the business of achieving competence in the LFC. Perceived as an obstacle to good pronunciation because of the negative impact L1 phonological transfer has on the target NS accent, the learner's mother-tongue phonology has traditionally been seen as an

impediment to successful learning. However, since an NS accent is not the goal for ELF, the value of the L1 phonology changes significantly, as we will see later in this chapter.

2. Improving accommodation skills
   With a whole chapter dedicated to this area in *The Phonology of English as an International Language*, Jenkins (2000) signaled the importance of accommodation skills from the outset. In the intervening years interest has continued to grow; Deterding underscored their importance when he stated that "the emphasis of the ELF proposals on developing accommodation skills … is exceptionally constructive and valuable for English language teaching" (Deterding 2011: 94).

   Depending on the levels of competence of the speaker(s)/listener(s) involved, accommodation will be need to be either receptive (with adjustments made to deal with incoming speech), productive (with adjustments made to the speaker's own pronunciation), or both. Jenkins (2000: 187–194) and Walker (2010: 88–92) described ways in which both receptive and productive accommodation skills can be taught in the classroom, and more recently Hancock has produced teaching materials addressing the same goal (Hancock 2012, 2013). These activities include:
   a. Student–student dictation. Learners dictate short texts to each other. In multilingual classrooms, this will expose learners to a range of English accents, which in itself is beneficial to them in terms of improved listening skills (Field 2003). In addition, if students are encouraged to seek/offer repetition of anything not fully understood, they gain experience in negotiating meaning. Moreover, the written record of a dictation allows teachers and learners to identify problem pronunciation items at an individual level.
   b. Communication activities. Less controlled than dictations, communication activities are an excellent way to improve accommodation skills for ELF. Whilst focusing on a communication task (guessing games, describe and draw, spot the difference, giving directions, etc.) students will tend to lapse into L1-influenced pronunciations, given the depth at which L1 phonological transfer operates. Where such transfer lies outside the LFC, meaning will not normally, in theory, be threatened. In contrast, erroneous pronunciation of items from the LFC should lead to communication breakdown. In their attempts to repair the "damage" in their communication, learners will need to converge on more target-like production of the item in question. In other words, whilst successful completion of a communication activity signals successful ELF pronunciation, failure to complete the task may signal that the pronunciation is problematical. At this point the teacher can intervene and help learners to identify the cause of the breakdown and determine if this lies in faulty production or faulty reception.

3. Working with monolingual groups
   The vast majority of ELT takes place in classes where learners, and very often their teachers, share the same L1. These classes cannot recreate the multilingual backgrounds found in language schools in the inner circle countries, and are necessary for the activities outlined in the previous section. As a result, it is fair to say that "much thought will have to be given to the problem of accommodation

Table 24.3 Communication tasks and pronunciation with multilingual and monolingual classes (from Walker 2010, adapted from Jenkins 2000).

Multilingual pair/group	→	Desire to communicate; convergence on common pronunciation	→	Replacement of unintelligible features from the mother tongue by items in the LFC	→	Intelligibility and reinforcement of items from the LFC
Monolingual pair/group	→	Desire to communicate; convergence on common pronunciation	→	Convergence on mother-tongue pronunciation	→	Intelligibility *but* reinforcement of mother-tongue accent

in groups containing members of the same L1" (Jenkins 2000: 193). This is because of the convergence on the L1 influenced forms that characterize attempts at increased intelligibility when interlocutors share a common mother tongue, as illustrated in Table 24.3.

In addition to the problem of L1 convergence, learners in a monolingual environment will receive only limited exposure to the range of accents that are commonplace in a multilingual class. This will reduce opportunities for learning to deal with accent variation and for needing to modify their own output.

In an examination of the reality of working on ELF pronunciation with monolingual groups, Walker (2001b) chose to focus on the benefits of such a situation in terms of both learner goals and the enhanced role of the NNS teacher. He later went on to suggest the use of student recordings as one way to ameliorate, although not resolve, the issue of convergence on the L1 phonology, and the subsequent divergence from internationally intelligible LFC forms (Walker 2005).

Improvements in receptive phonological accommodation are, fortunately, much less problematic for monolingual groups. Even though the group is geographically situated in a monolingual environment, technology today makes it easy for classes to access a multitude of ELF accents. Walker (2010: 95–96) mentions a number of websites that can be used to access accents from all over the world, exposure to which will help students to at least increase awareness of the problem of dealing with different accents.

The almost total absence of activities for accommodation skills training for monolingual groups stands in inverse proportion to the importance of such training for ELF users, who "must be prepared both to cope with major pronunciation differences in the speech of their different-L1 partners and to adjust their own pronunciation radically for the benefit of their different-L1 hearers" (Jenkins 2000: 194). In this respect, developing ways of improving phonological accommodation skills for such groups constitutes an important challenge for ELF pedagogy.

The learner's mother-tongue phonology

The *Contrastive Analysis Hypothesis* (CAH) (Lado 1957) held that second language (L2) phonology is filtered through the learner's first language. Similarity between the L2 and L1 phonologies is thought to lead to positive transfer, which equates with ease in the acquisition, whilst the difference leads to negative transfer and difficulty in acquisition. Negative transfer is commonly referred to as "interference", and while researchers today minimize the role that transfer plays in other areas of language acquisition, most agree that it operates strongly in L2 pronunciation. Given this, some approaches to teaching pronunciation for English as a foreign language openly make use of the L1 and L2 phonologies when determining priorities. In order to generate an inventory of phonological features for learners in a specific L1 context, for example, Brown (1992) suggests listing the phonemes and allophones of the L1 and L2 and determining the distributional restrictions on the L1 and L2 allophones and phonemes. Similarly, beginning her excellent summary of the processes of L1 phonological transfer (Jenkins 2000), Jenkins insists that a teaching syllabus for ELF "must be based on an understanding of the process of phonological transfer and its effects" (2000: 99).

There is a fundamental difference, however, between the value of the L1 in the teaching of pronunciation for English as a foreign language and its value for ELF. For the former, the learner's L1 is a root cause of error and L1 transfer is to be eliminated, or at least minimized, where it does not coincide with the NS target features. For the pronunciation of ELF this is not the case, since the goal here is not NS competence in the target features but intelligibility as determined by other NNSs, which, as we saw earlier, does not automatically equate with NS performance. This difference in goals allows us to view the learner's L1 phonology in a different light. Walker, for example, sees the learner's L1 as a resource rather than an obstacle and suggests that "[b]y openly starting from the learner's L1, we not only contemplate the reduced, achievable set of goals identified by the LFC. Equally importantly, we switch the emphasis towards what our learner CAN do (it is already part of their L1), and away from what they supposedly canNOT do" (2001b: 5). Jenkins puts the case for using the learner's L1 phonology more strongly: "[p]honological transfer is deep-rooted and can be of benefit to learners; it is not – and should not – be abandoned easily or willingly, unless there is very good reason to do so" (Jenkins 2000: 119).

There are two ways in which the L1 phonology can be of benefit to learners: the use of L1 allophones, accents, or related languages in order to achieve competence in target features of ELF pronunciation and the fine-tuning of the LFC to a specific L1 background. Both benefits can be optimized when the teaching takes place in a monolingual setting in which the learners have the same L1 background and in which the instructors are competent in the phonetics and phonologies of both English and the L1.

With regard to the first of these, Walker (2001b) demonstrated how allophones and L1 accents allow Spanish-L1 learners to achieve ELF-intelligible pronunciations

of a number of phonemes, including /z/, /ŋ/, /ʃ/, and /ʒ/: [z] is an allophone of /s/ in Spanish, occurring naturally in words like "asno" or "mismo"; [ŋ] is an allophone of /n/ before /k, g/, and is found in words like "banco" or "tengo"; /ʃ/ is a phoneme in various regional languages in Peninsular Spain; /ʒ/ is found in the "y" or "ll" of Argentinian Spanish, as in the words "yo" or "calle".

The approach works for other L1s. Arabic-L1 speakers, for example, have problems with /p/ in English, which often sounds like /b/. However, [p] exists in Arabic as an allophone of /b/ as in [sʌpt] ("Saturday") and [kʌpt] ("depression"). Since the allophonic variation of /b/ in this case is similar to English /p/, teachers can draw their students' attention to this /p/-like sound and then use this variant to train learners to pronounce a similar /p/ in their English.

Loan words are a related L1 resource that teachers can draw upon. Berger (in Walker 2010) suggests the use of *Adaggio* from Italian to help German-L1 speakers to pronounce /dʒ/. Similarly, in the Gulf states the Chinese origin loanword "chai" /tʃaɪ/ is commonly used when referring to tea, thus providing easy access to /tʃ/ for Arab-L1 learners from this area, whilst in Malay word-final /tʃ/ can be accessed through the loanword "Mac" (March), where the sound is word-final.

The second issue to consider with respect to the use of the learner's L1 phonology is the fine-tuning of the LFC. Jenkins (2000), for example, considers the quality of /ɜː/ a core feature; in her study, the Japanese speaker replaced /ɜː/ with /ɑː/ and was unintelligible to her interlocutor. However, Zoghbor (2011b) found that when Arab leaners replaced /ɜː/ by /eɪ/, this did not cause intelligibility problems, suggesting that the quality of /ɜː/ is a non-core feature for Arab learners. In contrast, an empirical study of Arab learners' word stress on words of more than two syllables (Zoghbor 2011b) suggests that this *is* a core feature for ELF intelligibility for this particular learner group.

Fine-tuning can also reveal gaps in the LFC. In a small-scale study of the intelligibility of Brazilian students, da Silva Sili (1999) found that the while the /r/-/h/ conflation typical of Brazilian speakers was problematic, the most common difficulty came from listeners not hearing or failing to identify the last syllable in words like "gazing", "happen", "patches", or "fancy". This led him to conclude that "the strong reduction of final syllable vowels by the speakers is not included by Jenkins in her 'core areas', but must definitely be considered a major area for error elimination in the speech of Brazilian students" (da Silva Sili 1999: 24).

Concerns regarding teaching pronunciation for ELF

Surveys and questionnaires have revealed that both learners and teachers harbor concerns about ELF pronunciation, whilst some linguists have accused ELF of leading to a lowering of standards. Keys and Walker (2002), Jenkins (2007: 22–28), and Walker (2010: 49–61) offer a full treatment of these and other concerns and misinterpretations, and what follows is restricted to learner and teacher preferences.

Learner preferences

A number of studies indicate that most learners want to sound like a native speaker. Dalton-Puffer, Kaltenboeck, and Smit (1997) surveyed 132 university students of English and asked them to rate unidentified native (RP, near RP, and GA) and non-native (Austrian) accents of English. The majority rated RP as their favorite model for pronunciation. The Austrian English accent was rated lowest, whilst in general the ratings reflected the respondents' familiarity with a given accent.

Timmis (2002) surveyed 400 teachers and students, exploring preferences for native-speaker norms not just for pronunciation but also for written and spoken grammar. In terms of pronunciation goals, two-thirds of his respondents showed a preference for native speaker competence. Grau (2005) asked first-year university students of English what the objective should be in German schools regarding pronunciation. Results showed that 65% opted for international intelligibility, as opposed to near-nativeness, but 59% then went on to say that neither /s/ nor /d/ were acceptable substitutions of the interdental fricative "th", despite the fact that Jenkins (2000) argued both variations are internationally intelligible.

Scales et al. (2006) analyzed the perceptions that 37 English language students and 10 NNS undergraduate students had of four accents: GA, BrE, Chinese English, and Mexican English. They found that "[w]hen asked to choose between wanting to be easily understood and having a native accent, the majority (62%) of English learners stated that their goal was to sound like a native speaker, compared with 38% who listed intelligibility as their pronunciation goal" (2006: 723). Interestingly, though, only 29% of the respondents were able to actually identify the American accent when asked to do so. In a blind listening task, a subject who had stated that her Asian classmates were less intelligible, chose the Chinese accent "as the easiest to understand and the one she liked most" (2006: 734). In general, Scales and her colleagues found the respondents to be unaware of the issues behind the choice of accent in a world where English has become the lingua franca.

Jenkins attributes some of the above contradictions to the *linguistic insecurity* that NNSs have as an outcome of the "negative stereotyping of their English by the NS community" (2004: 39). In this respect, it is interesting to note that in Timmis' study, where two-thirds of the respondents had shown a preference for an NS accent, this figure was actually reversed when students from outer circle countries were analyzed separately. This could be accounted for by the fact that outer circle countries are endonormative regarding English, and consequently possess greater *linguistic security* with respect to their own English accents and pronunciation. When Tokumoto and Shibata surveyed Japanese, Korean, and Malaysian university students, they found that while the Japanese and Koreans preferred an NS accent, the Malaysian students highly valued their accented English (Tokumoto and Shibata 2011).

What is clear at the present time is that the vast majority of learners know very little about English as a lingua franca. It may be that once they are "in full possession of the socio-linguistic facts" (Jenkins 2004: 36), and once teachers see ELF

intelligibility as a legitimate goal, learner preferences will shift towards ELF in those contexts in which ELF reflects the dominant use a group will make of their English.

Teacher preferences

The results of teacher surveys largely parallel those of learners. In a survey of teachers in Spain (Walker 1999), almost two-thirds said they would choose either RP or a standard British accent for their own pronunciation, with 75% selecting an NS accent. A study involving NS and NNS teachers working in Greece and the UK (Hannam 2006) found that "the majority of British participants were very critical of stigmatized British accents such as Liverpool and Belfast" (2006: 4) and would not use either in the classroom. In contrast, almost all of the Greek participants were positive about both accents, "with 100% saying they would use the Liverpool accent in the classroom and 75% the Belfast accent" (2006: 4). The Greek participants were much more critical of their own English accent, however, "with only 50% saying they would be happy to use this as a model" (2006: 4), whereas all of the British participants were positive about the idea of using the Greek English accent as a classroom model.

Jenkins carried out extensive research into teachers' attitudes to ELF in general and ELF accents in particular (2007), and found that with regard to NNS teacher preferences "NS accents, and particularly UK and US accents, [were] preferred in all respects by this large group of expanding circle respondents" (2007: 186).

Overall, English language teachers, especially NNS teachers, value NS accents highly. One explanation for this might be related to the prestige that an NS accent can give a teacher. Good teachers want to display very high levels of competence in the language they teach, grammatical, lexical, and phonological, and for the moment phonological competence is still seen in terms of proximity to a native-speaker accent. This argument is put forward by Wach (2011) on analyzing the results of a survey of 234 Polish students, who, as English majors, are destined to become teachers of English.

There are, of course, external restraints that condition the desirability and legitimacy of ELF accents. Although individual teachers may feel drawn to ELF, they could find it difficult to implement this desire in their classrooms. Colleagues might object on the basis that they are using a traditional NS accent as their model and so are worried that an ELF approach might confuse learners. Similar objections might also come from Directors of Studies or Principals. This is especially likely in private language schools, where marketing is frequently articulated around the employment of NS teachers and models.

These sorts of pressures help to explain why teachers who responded positively to the concept of ELF accents at a theoretical level, "did not think that it would be feasible to implement the teaching of ELF accents in classrooms in their own countries or even, in most cases, to use their own proficient NNS accents as pronunciation models" (Jenkins 2007: 224).

The situation is further complicated by the fact that many international exam boards assess pronunciation in terms of the presence or absence of a foreign accent.

Until this changes, teachers preparing learners for such exams are obliged to take NS accents into account. It is possible, however, that the overall attitude to ELF will change in the future. Referring to the situation for assistant English teachers in Japan, Sutherland (2008) suggests that "as awareness of ELF increases, students, their parents and other interested parties will realize that Japanese teachers should not be characterized as NNESs, with all the negative associations implied by that term, but should instead be seen as proficient ELF speakers" (2008: 10). In the meantime, teachers can become agents for raising awareness in their local environment, beginning with their colleagues, Directors of Studies, and Principals.

The teaching of the pronunciation of English as a lingua franca is a complex business, and this chapter has only been able to provide a brief first contact. For some, this will be a first and last contact, since ELF is a subject that generates sometimes fierce opposition, something the authors are fully aware of. However, everything points to English being the world's leading lingua franca for some time to come, and on a daily basis anyone operating in this brave new world will come across examples of successful spoken communication despite decidedly non-standard, non-native speaker accents. How can this be? How can communication succeed with pronunciations so far removed from the native-speaker norm? And yet it does. Research into ELF pronunciation attempts to understand *how* it does and how to convert these findings into coherent pedagogy.

REFERENCES

Avery, P. and Erhlich, S. 1992. *Teaching American English Pronunciation*, Oxford: Oxford University Press.

Benrabah, M. 1997. Word stress – a source of unintelligibility in English. *International Review of Applied Linguistics in Language Teaching* 35(3): 157–165.

Brown, A. 1992. *Approaches to Pronunciation Teaching*, London: Modern English.

Cauldwell, R. 2006. Brazil's yes-no questions. Message posted to supras@mailbox.gsu.edu on May 18, 2006.

Dalton, C. and Seidlhofer, B. 1994. *Pronunciation*, Oxford: Oxford University Press.

Dalton-Puffer, C., Kaltenboeck, G., and Smit, U. 1997. Learner attitudes and L2 pronunciation in Austria. *World Englishes* 16: 115–128.

da Silva Sili, R. 1999. A small-scale investigation into the intelligibility of the pronunciation of Brazilian intermediate students. *Speak Out! Newsletter of the IATEFL Pronunciation Special Interest Group* 23: 19–25.

Derwing, T. 2006. How pronunciation research can inform teaching practices. In: *Proceedings of the 15th International Symposium on English Teaching*, Y. Chen and Y. Leung (eds.), Taipei, Taiwan: English Teachers' Association.

Derwing, T.M. and Munro, M.J. 1997. Accent, intelligibility, and comprehensibility: evidence from four L1s. *Studies in Second Language Acquisition* 19: 1–16.

Derwing, T.M. and Munro, M.J. 2008. Putting accent in its place: rethinking obstacles to communication. Retrieved from: http://www.sfu.ca/~mjmunro/D&MHandout.pdf.

Deterding, D. 2010. ELF-based pronunciation teaching in China. *Chinese Journal of Applied Linguistics* 33(6): 3–15.

Deterding, D. 2011. English language teaching and the Lingua Franca Core in East Asia. In: *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS XVII)*: vol. 95. Retrieved from: http://www.icphs2011.hk/ICPHS_CongressProceedings.htm.

Field, J. 2003. The fuzzy notion of "intelligibility": a headache for pronunciation teachers and oral testers. *IATEFL Special Interest Groups' Newsletter in Memory of Gill Porter Ladousse:* 34–38.

Giles, H. and Coupland, N. 1991. *Language: Contexts and Consequences*, Milton Keynes: Open University.

Giles, H., Coupland, J., and Coupland, N. 1991. Accommodation theory: communication, context and consequences. In: *Context of Accommodation*, H. Giles, J. Coupland, and N. Coupland (eds.), Cambridge: Cambridge University Press.

Grau, M. 2005. English as a global language – What do future teachers have to say? In: *The Globalization of English and the English Language Classroom*, C. Gnutzmann and F. Intemann (eds.), Tübingen: Gunter Narr.

Hancock, M. 2012. At the talk face 8: Accommodation games 1. *Speak Out! Newsletter of the IATEFL Pronunciation Special Interest Group* 47: 30–31.

Hancock, M. 2013. At the talk face 9: Accommodation games 2. *Speak Out! Newsletter of the IATEFL Pronunciation Special Interest Group* 48: 30–31.

Hannam, S. 2006. Pronunciation teaching today: listening without prejudice. *TESOL-SPAIN Newsletter* 29: 3–6.

Jenkins, J. 1998. Pronunciation norms and models. *ELT Journal* 52(2): 119–126.

Jenkins, J. 2000. *The Phonology of English as a Lingua Franca*, Oxford: Oxford University Press.

Jenkins, J. 2004. ELF at the gate: the position of English as a lingua franca. In: *IATEFL 2004 Liverpool Conference Selections*, A. Pulverness (ed.), Canterbury: IATEFL.

Jenkins, J. 2005. Teaching pronunciation for English as a lingua franca: a sociopolitical perspective. In: *The Globalization of English and the English Language Classroom*, C. Gnutzmann and F. Intemann (eds.), Tübingen: Gunter Narr.

Jenkins, J. 2007. *English as a Lingua Franca: Attitude and Identity*, Oxford: Oxford University Press.

Kachru, B. 1985. Standards, codification and sociolinguistic realism: the English language in the outer circle. In: *English in the World: Teaching and Learning the Language and Literature*, R. Quirk and H. Widdowson (eds.), Cambridge: Cambridge University Press.

Kharma, N. and Hajjaj, A. 1997. *Errors in English among Arabic Speakers: Analysis and Remedy*, Beirut: York Press.

Kenworthy, J. 1987. *Teaching English Pronunciation*, Harlow: Longman.

Keys, K. and Walker, R. 2002. Ten questions on the phonology of English as an international language. *ELT Journal* 56(3): 298–302.

Lado, R. 1957. *Linguistics across Cultures: Applied Linguistics for Language Teachers*, Ann Arbor, MI: University of Michigan Press.

Marks, J. 1999. Is stress-timing real? *ELT Journal* 53(3): 191–199.

Matsuura, H., Chiba, R., and Fujieda, M. 1999. Intelligibility and comprehensibility of American and Irish Englishes in Japan. *World Englishes* 18: 49–62.

Munro, M.J. and Derwing, T.M. 1995. Foreign accent, comprehensibility and intelligibility in the speech of second language learners. *Language Learning* 45: 73–97.

Nasr, R.T. 1963. *The Teaching of English to Arab Students*, London: Longman.

Nelson, C. 2008. Intelligibility since 1969. *World Englishes* 27(3): 297–308.

Osimk, R. 2009. Decoding sounds: an experimental approach to intelligibility in ELF. *Vienna English Working Papers* 18(1): 64–89. Retrieved from: http://anglistik.univie.ac.at/fileadmin/user_upload/dep_anglist/weitere_Uploads/Views/0901final.pdf.

Pennington, M. 1996. *Phonology in English Language Teaching: An International Approach*, London: Longman.

Rajadurai, J. 2006. Pronunciation issues in non-native contexts: a Malaysian case study. *Malaysian Journal of ELT Research* 2: 42–59.

Roach, P. 1991. *English Phonetics and Phonology*, 2nd edition, Cambridge: Cambridge University Press.

Roach, P. 2009. *English Phonetics and Phonology*, 4th edition, Cambridge: Cambridge University Press.

Scales, J., Wennerstrom, A., Richard, D., and Hui Wu, S. 2006. Language learners' perceptions of accent. *TESOL Quarterly* 40(4): 715–738.

Seidlhofer, B. 2005. English as a lingua franca. *ELT Journal* 59(4): 339–341.

Seidlhofer, B. 2011. *Understanding English as a Lingua Franca*. Oxford: Oxford University Press.

Smith, L.E. and Nelson, C. 1985. International intelligibility of English: directions and resources. *World Englishes* 4(3): 333–342.

Sutherland, S. 2008. Team teaching with NES and NNES teachers. *IATEFL Voices* 205: 10.

Timmis, I. 2002. Native speaker norms and international English. *ELT Journal* 56(3): 240–249.

Tokumoto, M. and Shibata, M. 2011. Asian varieties of English: attitudes towards pronunciation. *World Englishes* 30(3): 392–408.

Trudgill, P. 1999. Standard English: what it isn't. In: *Standard English – The Widening Debate*, T. Bex and R.J. Watts (eds.), London: Routledge.

Ur, P. 2010. English as a lingua franca: a teacher's perspective. *Cadernos de Letras (UFRJ)* 27 (December 2010): 87. Retrieved from: http://www.letras.ufrj.br/anglo_germanicas/cadernos/numeros/122010/textos/cl301220100penny.pdf.

Wach, A. 2011. Native-speaker and English as a lingua franca pronunciation norms: English majors' views. *Studies in Second Language Learning and Teaching* 1(2). Retrieved from: https://repozytorium.amu.edu.pl/jspui/bitstream/10593/2135/1/SSLLT%201(2)%20247-266%20Wach.pdf.

Walker, R. 1999. Proclaimed and perceived wants and needs among Spanish teachers of English. *Speak Out! Newsletter of the IATEFL Pronunciation Special Interest Group* 24: 25–32.

Walker, R. 2001a. International intelligibility. *English Teaching Professional* 21: 10–13.

Walker, R. 2001b. Pronunciation priorities, the Lingua Franca Core, and monolingual groups. *Speak Out! Newsletter of the IATEFL Pronunciation Special Interest Group* 28: 4–9.

Walker, R. 2005. Using student-produced recordings with monolingual groups to provide effective, individualized pronunciation practice. *TESOL Quarterly* 39(3): 535–542.

Walker, R. 2010. *Teaching the Pronunciation of English as a Lingua Franca*, Oxford: Oxford University Press.

Weinberger, S. 1987. The influence of linguistic context on syllable simplification. In; *Interlanguage Phonology: The Acquisition of Second Language Sound System*, G. Ioup and S. Winberger (eds.), Cambridge: Newbury House Publishers.

Zawaydeh, B., Tajima, K., and Kitahara, M. 2002. Discovering Arabic rhythm through a speech cycling task. In: *Perspectives on Arabic Linguistics*, D. Parkinson and E. Benmamoun (eds.), Amsterdam: Benjamins.

Zielinski, B. 2004. Measurement of speech intelligibility: What are we actually measuring? Paper presented at the annual meeting of the American Association for Applied Linguistics, Portland, OR. In: Yang, J. 2009. *English Phonological Intelligibility: A Phono-Numerical Measure*. Retrieved from: http://teacher.yuntech.edu.tw/yanght/cv/phono-measure-reference.pdf.

Zoghbor, W. 2011a. Teaching the pronunciation of English as a lingua franca: reducing skepticism and increasing practicality. *International Journal of Humanities and Social Science* 1(8): 284–288.

Zoghbor, W. 2011b. The effectiveness of the Lingua Franca Core (LFC) in improving the perceived intelligibility and perceived comprehensibility of Arab learners at post-secondaryl. Unpublished doctoral thesis, Leicester, UK: The University of Leicester.

25 Intonation in Research and Practice: The Importance of Metacognition

MARNIE REED AND
CHRISTINA MICHAUD

Introduction

Intonation, as defined by Pickering (2012), is "the systematic and linguistically meaningful use of pitch movement at the phrasal or suprasegmental level" (2012: 280). In 1999, John Levis analyzed the teaching of intonation and argued that "present intonational research is almost completely divorced from modern language teaching and is rarely reflected in teaching materials" (1999: 37). In the years since Levis made this claim, the field has continued to advance, giving reasons for optimism regarding the convergence of research and teaching materials. However, for a variety of reasons, intonation remains a challenge for teachers and students alike, at both the metacognitive and skill levels.

Although excellent suprasegmental textbooks exist, with sections on intonation informed by research, these often focus on getting learners to produce the target intonation itself. Nevertheless, teaching intonation must include metacognitive awareness as well as productive and receptive skills if it is to be successful.

This chapter will consider pedagogical approaches to intonation in theory and in practice, using both examples from textbooks and data from an original study looking at intonation attitudes of L2 learners.

The missing link between theory and practice seems to be metacognition. Citing Goh (2008), Vandergrift and Goh (2012) state: "Metacognition refers to listener awareness of the cognitive processes involved in comprehension, and the capacity to oversee, regulate, and direct these processes" (2012: 23). Bringing research on intonation and metacognition into the classroom has not happened with great consistency, though some work, focusing mainly on the concept of intelligibility in listening and speaking, has begun to look at the role of metacognition and strategy

The Handbook of English Pronunciation, First Edition. Edited by Marnie Reed and John M. Levis.

instruction and has implications for the study of intonation (Mendelsohn 1998; Chamot 1995; Goh 2008; Vandergrift and Goh 2012; Rost 2005).

We conclude this chapter with specific recommendations for better classroom practice when teaching intonation with a metacognitive focus.

Theories informing intonation pedagogy

Although it is possible to address intonation in English from different perspectives (see Levis and Wichmann, Chapter 8 and Wichmann, Chapter 10 in this volume), for the purposes of teaching intonation to L2 learners we are primarily interested in the key role intonation plays in implicature.

An overview of research into the treatment of intonation by phonologists and pragmatists is provided by Wharton (2012), who situates the relationship between intonation and inferential intentional communication within a Gricean framework. Wells (2006) goes further in his investigation into the implicational use of intonation, identifying what he calls "the implicational fall-rise", when a "speaker *implies* something without necessarily putting it into words[….] By making a statement with the fall-rise, the speaker typically states one thing but implies something further. Something is left unsaid – perhaps some kind of reservation or implication" (2006: 27). Studies of intonation in English as an L2 framed from the perspective of speech act theory (Searle 1969) support the view that this implicational function of intonation is key by including stress and intonation contours (Searle and Vanderveken 1985) as among the devices helping to draw learners' attention to the illocutionary focus of an utterance.

An open question is the ability of even advanced L2 listeners to attend to prosodic cues or credit intonation with "the power to reinforce, mitigate, or even undermine the words spoken" (Wichmann 2005: 229). Intonation is certainly much more important for ESL than EFL or ELF contexts, where other means than intonation will often be used to indicate stress, focus, and speaker intent (see Hirst and DiCristo 1998). However, when teaching English in settings where non-native speakers (NNSs) will interact frequently with native speakers (NSs) of English, intonational implicature becomes an essential component of instruction, since NSs often use the implicational fall-rise unconsciously; therefore it is less likely to be taught explicitly and consequently unlikely to be attended to by NNSs.

Intonation in practice: an overview of current approaches and relevant research

Teaching materials and textbooks

Intonation is currently addressed in many teacher reference books with significant pronunciation components (Brown 2011; Celce-Murcia et al. 2010; Grant 2014) as well as in what Murphy and Baker (Chapter 3 in this volume) refer to as "Activity

Recipe Collections". Excellent sections on intonation, informed by research, are also present in popular suprasegmental pronunciation textbooks available for stand-alone courses or for use as supplemental resources in regular ESL classes (Gilbert 2012; Grant 2010; Hahn and Dickerson 1999; Miller 2006). Many integrated skills or general speaking/listening textbooks also include sections on intonation as well, the most frequently cited in a recent survey (Foote, Holtby, and Derwing 2011) being *Side by Side* (Molinsky and Bliss 2002), a course book that systematically integrates pronunciation skills.

Three main aspects of intonation are treated in these texts: (1) intonation contours over phrases and sentences, resulting in sentence-final pitch changes; (2) intonation signaling attitudes and emotions; and (3) intonation accompanying changes in phrase or sentence focus (sometimes also called sentence stress). We will consider each of these with examples from textbooks.

Intonation contours and sentence-final pitch changes Intonation contours over the length of a sentence or question are one major aspect of intonation that is taught to ESL students. Learners are introduced to intonation fairly early on in grammar-based, integrated skills, or listening/speaking classes in this manner, likely because the known grammar can help scaffold the new intonation. One example from the *Listening and Speaking* 1 volume of the widely used *Tapestry* series is typical of this approach to intonation. After beginning by asking learners to distinguish general "falling" intonation in sentences from general "rising" intonation in questions, the book continues:

> Listen carefully as your teacher asks these questions:
>
> What are you going to do after class?
> Are you going to study after class?
>
> Does his or her voice sound different at the end of each question? When you ask an information question (a question that begins with *who*, *what*, *when*, *where*, or *how*), the tone of your voice usually rises a little at the end of the question. When you ask a yes/no question, the tone of your voice goes down at the end of a question. (Benz and Dworak 2000: 247)

Learners are then faced with a long list of questions, both wh questions and yes/no questions, and are instructed to read them aloud and focus on the final intonation. In later sections and at higher levels, learners are also introduced to the final intonations of tag questions and either/or questions in this same manner. This approach is typical of "textbooks [that] have presented elaborate technical rules for intonation … based on grammar" (Gilbert 2014: 113). When surveying the field, Levis (1999: 48), citing others, found that "Even textbooks that eventually give a more complete view start with this kind of rule." Nevertheless, this approach often leads learners to produce (at least initially) questions or sentences with exaggerated and unnatural final intonation. Since we know that even experts may disagree on speaker intent when analyzing the "correct" pitch contours of

different samples (Lieberman 1967: 124), learners may similarly make predictions about English that are not completely supported by evidence.

Emotional and affective elements of intonation Another way that intonation is often taught to learners is by referencing its role signaling emotion and speaker attitude. Linda Grant (2010), in her pronunciation text, *Well Said*, introduces this affective aspect of intonation by having learners listen to a two-line dialogue on the accompanying CD:

> In this example, how does speaker Y indicate surprise?
>
> *Example:* X: He has 10 brothers.
> Y: He has 10 brothers? (I'm really surprised.)
>
> *You can show surprise or disbelief by using rising pitch to echo a statement. The pitch rise is usually on the stressed syllable of the last content word.* (2010: 113)

This approach to intonation instruction, in contrast to the grammatical approach, seeks to engage learners in mimicking the exaggerated prosody of English, such as the large pitch variations (Collier and Hubbard 2001) associated with emotional states such as happiness and (as in this example) surprise. Aided by authentic audio or video clips, learners practice producing intonation contours and identifying the underlying speaker affect, including differentiating sincerity from sarcasm. Empirical support certainly exists for an approach that encourages learners to be sensitive to the use of intonation to convey speaker attitude and emotion. As noted by Gumperz (1982) in his seminal cross-cultural examination of the extent to which intonation determines how a speaker's message is understood, non-native intonation may result in negative social evaluation.

There are drawbacks, however, to an exclusive pedagogical focus on identifying and expressing attitudinal and emotional aspects of intonation. When making decontextualized judgments, including judgments of sincerity or sarcasm, differences and disagreements have been reported between speakers' intended meaning and listeners' interpretations (Beun 1990; Uldall 1964). In addition to the subjective nature of these judgments, sarcasm is a late acquisition in L1 English (Berko Gleason and Ratner 2009) and therefore might be problematic as the basis for early teaching of intonation in L2 English.

While an emphasis on the emotional side of intonation can be taken to extremes, what makes Grant's example, above, successful is the accompanying explanation, which guides learners to focus on the function of these pitch changes.

Intonation and focus or stress within phrases and sentences As noted by Couper-Kuhlen (2001), "intonation – in the restricted sense of 'pitch configuration' – rarely functions alone to cue an interpretive frame" and should be considered in conjunction with other prosodic phenomena including timing and volume (2001: 16). In practical terms, teaching intonation often means considering final intonation in conjunction with sentence focus.

Texts typically introduce sentence focus in the context of given and new information, explaining that speakers use rising pitch on content words (versus function or structure words) but then also on new information (versus old, or given, information):

> New information refers to words or ideas in a message unit that are new to the conversation. They are words not used before or ideas not already obvious to the speakers. New information is often found at the end of a message unit (Hahn and Dickerson 1999: 63).
>
> A: What kind of triangle is this?
> B: It's a right triangle. (1999: 64)

Exercises then follow that ask learners to mark the new information in a conversation or passage and practise reading it aloud with rising pitch on the new information.

Textbooks often then move on to showing learners that in English speakers can choose to stress any word in an utterance with a different intended meaning. For example:

> (a) He CALLED yesterday.
> (b) HE called yesterday.
> (c) He called YESTERDAY.
>
> Stress on different words can change the meaning of a sentence. In (a) the emphasis is on **called**, rather than another action, such as coming in person. In (b), **he**, instead of someone else, called. In (c), he called **yesterday**, not another day. (Hagen 2000: 118)

Though this particular textbook (and many like it) describes what is happening in these utterances as changes in stress, we note that stress and intonation, in this case, are inextricably linked. However, while English speakers certainly have the option to use marked stress and intonation to encode pragmatic function and signal alternate meanings (i.e., make implications), we also have the option of varying our syntax:

> (a) What he did was call yesterday.
> (b) He was the person who called yesterday.
> (c) It was yesterday that he called.

While every language has at least one mechanism for signaling the "point of information focus" (Bolinger 1972), L2 learners whose L1s use only morphosyntactic mechanisms are not generally used to relying on intonation to help decode the meaning of the message. This suggests that many ESL learners may not notice the role of intonation in communicating speaker intent (Pennington and Ellis 2000) and may instead be relying on their native language's default mechanism, which is often syntactic or lexical rather than intonational.

Research suggests that native speaker listeners rely heavily on the combination of final intonation and focus in utterances to make sense of larger discourse (Hahn 2004). When speakers misplace focus in a sentence or do not use intonation and focus to signal appropriate contrasts and given-new information statuses, native speaker listeners find it harder to follow the message:

> The urban environment is more individualistic than the rural environment [expected given-new stress and intonation].
>
> The urban environment is more individualistic than the rural environment [unexpected given-new stress and intonation]. (2004: 206)

Therefore, the role of intonation and sentence focus is essential for interpreting speaker intent. Beyond merely telling whether a speaker is surprised or not, listeners need to be able to make inferences on the basis of the speaker's intonation signal. The second edition (1993) of Judy Gilbert's *Clear Speech* has a useful introduction for students to this concept:

> You can often guess what will come next by noticing which word the speaker has emphasized. Guessing what will come next is a good way to listen to English more effectively. (1993: 90)
>
> a. We prefer beef soup. Not stew?
> b. We prefer beef soup. Not chicken? (1993: 91)

Exercises like this exist in all the other editions of *Clear Speech*, as well as in other pronunciation texts, and can certainly help focus learners on particular stressed words in a given utterance. However, the explicit introduction to the concept of inferencing ("guessing") based on intonation and stress – the underlying idea that informs all exercises of this type – is important and is sometimes presented only implicitly in other texts.

Original research on intonation

In a pilot study, Reed investigated learner listening skill for and metacognitive awareness of the pragmatic function of intonation to signal speaker intent. Data were gathered in two intact pronunciation elective classes in an academically oriented intensive English program on a university campus in the Northeast. Subjects were high-intermediate and advanced-level students receiving segmental and suprasegmental instruction from seasoned instructors, both using the same pronunciation course book with a prosodic focus. Students received instruction in stress and intonation, including explicit instruction and lab practice producing marked intonation contours and contrastive stress. The researcher sat in on every class session and administered additional diagnostics and assessments at three points in the semester. Pre-instruction assessments of learners' perceptual awareness and metacognitive beliefs about English intonation were administered.

Anonymous student response systems (clickers) were used to elicit multiple choice and true/false responses to determine students' beliefs regarding the functions of intonation. Finally, students' abilities to make inferences were assessed aurally using two recorded sentences, one with unmarked and one with contrastive stress, requiring forced-choice responses.

Pre-instruction student responses revealed robust perceptual awareness of English stress and intonation. To determine whether learners perceptually noticed the rhythmic characteristics of English they were asked to identify which, if any, of three one-minute speech samples "sounded like" English, and to report the basis for their determination. The samples, one each from English, French, and Japanese (arguably representing a stress-timed, syllable-timed, and mora-timed language respectively) were same-topic NPR, Le Monde, and NHK radio news broadcasts that had been filtered (low-pass, 400 Hz) to remove lexical information usable to distinguish the languages. Learners accurately identified the English sample, but expressed negative perceptions of English intonation. Specifically, though the English sample contained unmarked (normal) intonation, learners dismissed it as "exaggerated" and noted the "sing-songy" pitch contours as the mechanism by which they distinguished English from the other two languages.

One finding of the pretest was that learners did not attend to marked intonation and sentence focus when trying to interpret an utterance. This was consistent with what Pennington and Ellis (2000) found. The pre-test included the following example:

> The *teacher* didn't grade your papers.

When asked to answer the question "Have the papers been graded?" learners initially responded in the negative. Told that the answer was in fact positive, they then asked for repeated hearings of the audio recording and mouthed the words "didn't grade" to themselves while listening. Their responses to the question indicated that they did not believe that intonation had the ability to override words; 70% of the learners in one class and 100% of the learners in another ($N = 14$ in each class) replied "No" incorrectly, simply because they did not attend to the signal of the marked intonation.

Post-instruction teacher surveys revealed instructor satisfaction on having successfully taught stress and intonation, as measured by students' coached language-lab production, which did converge on the target intonation of English.

> Example: "*Some* companies in the high-tech sector sell a wide variety of products."

Nevertheless, despite their awareness of the general intonation contours of English and their successful production of the marked intonation, learners were unable to discern the underlying meaning (implication) signaled by marked intonation. In the above example, when asked what the speaker would go on to discuss, learners said the variety of products, referencing sentence position. While NS listeners might predict that the next sentence would discuss other companies,

no learners picked up on the implication signaled by the very same marked intonation they had practised the week before. One student explicitly questioned that idea and said, "If this [intonation] was really important, someone would have told us by now."

Learners' strategies for listening did not change over the course of the semester and remained consistent with their beliefs (which also did not change despite the production-focused instruction) that intonation is unimportant and that words trump intonation. Both pre- and post-instruction, learners expressly rejected a role for intonation in overriding surface lexical information; maintained that the sole mechanism for conveying meaning is through the locution, the words of the utterance; and were unable to use intonation, when listening, to grasp speaker intent, the illocution.

Post-instruction surveys conducted by the researcher revealed continuing learner uncertainty about the real-life applications or significance of intonation. Students rejected ever voluntarily producing these patterns outside the classroom, stating they felt "foolish" when producing the target intonation and that the patterns sounded "silly" and "ridiculous".

Discussion: research and practice divides

In this study, one of the underlying questions that emerged is how to gauge when learners have truly "learned" intonation. The teachers in the study progressed through the materials in the book, which are cumulative and communicative in nature; one expressly commented that students had "learned" intonation after she taught it and they in fact produced it. However, with students finishing the class rejecting the entire idea of marked intonation – both for listening as well as for their speaking – it seems problematic to say that they have actually learned anything about intonation other than the ability to mimic it. Instruction did not move beyond a productive level to a metacognitive level, and because only the researcher asked questions about learners' strategies and metacognitive beliefs, the teachers were not aware of a problem.

This finding is echoed by others in the field. Gilbert (2014) notes that "because the system [of English intonation] is apt to be foreign to students, they may not actually believe that intonation affects meaning" (2014: 125). She goes on to observe that learners "will rarely tell the teacher that they feel silly speaking this way, and the result will be that they may walk out of the class without having accepted the system at all. Or they may think intonation is simply decorative" (2014: 125).

In production-focused classrooms, therefore, learners may well produce the intonation contours on demand, but they may finish the course expressing uncertainty about the real-life applications or significance of these intonation patterns and expressing ambivalence about adopting the intonation patterns in their own speech outside the classroom (see Mennen, Schaeffler, and Doherty 2012). As observed by Paunović and Savić (2008), "Students often do not have a clear idea of why exactly 'the melody of speech' should be important for

communication, and therefore seem to lack the motivation to master it, while teachers do not seem to be theoretically or practically well-equipped to explain and illustrate its significance" (2008: 72–73). While current research on intelligibility and intonation in general has moved beyond the native-speaker model (Levis 2005) and acknowledges that some aspects of pronunciation may not be relevant to contexts in which NNSs communicate only with other NNSs (Jenkins 2000, 2002), learners nevertheless need to be able to draw on their understanding of intonation and its pragmatic functions in order to make sense of the implicational fall–rise patterns that NSs use. As Tomlinson and Bott (2013) state, "often what a speaker intends to say is not always directly retrievable from a linguistic form; rather listeners must *infer* it" (2013: 3569). Therefore, NNS perception is crucial, and so is the ability to not only hear but also interpret marked intonation in English.

To summarize, a narrow focus on production in suprasegmental instruction may lead teachers to falsely assume that students have "learned" intonation and contrastive stress. Teachers may be unaware that students may not only be unwilling to use these patterns in their own speech but also be unaware of the role of intonation in signaling speaker intent. Therefore, production-focused instruction, without an overtly metacognitive approach, masks a gap in instructor and student (meta)cognition.

In part, this gap exists because teachers themselves may have had limited training in teaching intonation. In their survey of pronunciation teaching practices in Canada, Breitkreutz, Derwing, and Rossiter (2001) found that only 30% of surveyed teachers had received any kind of training in pronunciation. A follow-up study ten years later by Foote, Holtby, and Derwing (2011) found that, "For the most part, instruction in pronunciation in Canada has not changed in the last decade" (2011: 1). Since much of this instruction can be assumed to be segmental in nature, we hypothesize that far fewer than 30% of teachers, therefore, have received training in how to teach any of the suprasegmentals, including intonation.

Furthermore, intonation is acquired so early in L1 that it becomes ingrained to the extent that untrained NS teachers tend not to be aware of their own uses of it. We know that intonation (along with rhythm and other prosodic features) is one of the first aspects of an L1 acquired (DeCasper and Fifer 1980; DeCasper and Spence 1986; Spence and DeCasper 1987; Vihman, Chapter 19 in this volume). Newborn preference studies (Moon, Cooper, and Fifer 1993) reveal neonate attention to and preference for "the rhythms and sounds of language" including intonation, to which the infant has been exposed *in utero* (Karmiloff and Karmiloff-Smith 2001: 43). As Linda Grant has noted, "native speakers use suprasegmental features unconsciously. Like their students, native-speaking teachers are seldom aware of speech features like English rhythm and intonation and how they impact meaning unless those concepts are explicitly pointed out" (Grant 2014: 13–14).

We can find many examples of these subconscious uses of intonation in classrooms. In studies looking at types of teacher corrective feedback and their effectiveness, Lyster and Ranta (1997) found that intonation plays a key role in

corrective feedback containing a repetition "of the student's erroneous utterance. In most cases, teachers adjust their intonation so as to highlight the error (1997: 48). In a classroom setting, after repeated work on third-person singular present tense verb endings, a learner reported about the absence of one of his classmates, saying, "Teacher, every Friday Luis go to the bank." The teacher tried to point out the learner's error: "Luis *go* to the bank?" The teacher's stress and rising pitch on "go" here would have signaled to NS listeners the exact location of the error, but the learner in question did not attend to the intonation, and instead began attempting to repair his utterance by varying the preposition. The learner's failure to notice the focus in the teacher's utterance (signaled by the intonation) is not uncommon: Lyster and Ranta (1997) found that this type of repetition with pitch changes results in successful repair on the part of learners only 31% of the time.

In another classroom setting, when collecting essays on the day they were due, a teacher paused in front of a learner who did not have her essay. "Can I give it to you on Monday?" the learner asked. "You *can*," the teacher replied, implicitly indicating a "but" which was unstated (Wells 2006), in this case referring to the late penalty for papers listed on the syllabus. "Okay, thanks!" the learner replied with relief.

As we have seen with the examples from intonation sections in textbooks above, the field is moving toward a more explicit and metacognitive focus that will guide learners toward realizing the importance of these patterns. Nevertheless, teachers, such as the teachers in the study described above and those Grant (2014) mentions, may find it difficult to maximize the potential of such materials. In the absence of formal training in their graduate work, student textbooks therefore have become the de facto training mechanism for many teachers. More explicit statements about the implicational function of intonation, therefore, such as that included in the excerpt from Gilbert (1993), can help teachers as well as learners in the classroom.

Teachers need to be able to identify the mechanisms by which English signals contrast and/or implication and realize that these mechanisms are not linguistically universal. In the examples given above, teachers seem unable to suppress the innate and intuitive use of intonation for implication, even in the classroom, and even when talking to learners, and thus teachers can be said to simply not grasp what students do not grasp about intonation.

When teaching intonation, for example, it is logical to assume that teachers tend to go first to the topics in intonation that they are aware of themselves consciously manipulating (such as sarcasm, etc.); these aspects of intonation, along with the grammar-based intonation contours of sentences and questions, are indeed treated fully in many texts. Consequently, teachers may spend less time on the vast world of intonation that they use subconsciously, especially the implicational fall–rise.

Even for teachers who have not been trained in teaching pronunciation, the importance of intonation for their learners can be explained via reference to pragmatics. Teachers are used to explicitly teaching certain aspects of English pragmatics to learners – for example, teaching beginning-level learners in an ESL context that "Hi, how are you?" is not generally an invitation for them to tell the speaker how they actually are feeling that day. Along these same lines, such explicit

instruction into the pragmatic norms surrounding intonation (telling learners directly that intonation can trump the words in an utterance and signal specific alternate meanings) is essential.

Implications for a metacognitive approach to the classroom

As we have argued, relying solely on the production-driven side of intonation for instruction carries significant negative pedagogical consequences. As Grant (2014) argues about prosody in general, and we would argue about intonation in specific, "If the communicative value of suprasegmentals is not made clear, learners may decide learning suprasegmental features is not worth the effort" (2014: 19). This was true of the students reported in this chapter. Clearly, it is essential to go beyond the traditional focus on pronunciation alone, as Grant (2014) says: "As important as what happens on the perceptual, motor, and cognitive levels" in intonation instruction "are the conscious and unconscious attitudes of adult learners toward pronunciation change" (2014: 29). This "making clear" and these "conscious and unconscious attitudes" are of course metacognition in action.

Five recommendations for a metacognitive-focused approach to intonation

Label intonation patterns in English (marked versus unmarked) to aid learner metacognition

Learners need language to distinguish and describe different kinds of intonation patterns and articulate and discuss their underlying beliefs about intonation. In this chapter, we have been describing intonation as "unmarked" or "marked", but in a classroom, teachers may want to use more learner-friendly terms; we suggest using "normal intonation" and "special intonation". We could also imagine labeling unmarked intonation as neutral intonation or expected intonation, and marked intonation as signaling intonation or unexpected intonation.

In the original data reported earlier in this chapter, we noted that learners who initially were able to recognize unmarked English intonation perceived it negatively and as "exaggerated". Learners of course do not need to adopt normal English intonation into their own speech, though they may find certain advantages to doing so. However, they must be able to recognize that what they perceive as "exaggerated" intonation is in fact unmarked, normal English intonation; this change in their underlying attitudes is necessary in order to be able to perceive truly exaggerated (marked) intonation and realize its pragmatic functions. As Reed was solely a non-participant observer, not the classroom teacher in the study reported, learners were never given language to redefine "exaggerated" intonation as normal intonation, which could function as the first step toward evolving learner beliefs.

Use metacognitive diagnostics and assessments to frame instruction

Metacognition begins before teachers even start teaching, with an understanding of students' initial beliefs about pronunciation. Teachers may want to consider an initial diagnostic that focuses not only on students' intonation skills but also on their metacognitive grasp of intonation and its functions. An initial diagnostic could include, as well as items that test the production and perception of various aspects of intonation, items such as the following (Reed and Michaud 2005):

True or false? Intonation, stress, and timing can . . .

a. ☐ turn a statement into a question
b. ☐ turn a sincere statement into a sarcastic one
c. ☐ act as oral punctuation, quotation marks, and paragraph breaks
d. ☐ signal an implied contrast
e. ☐ change the meaning of a sentence
f. ☐ reduce the number of words needed to convey your meaning
g. ☐ convey information without actually saying the words.

Beginning a course or unit on intonation with a questionnaire like this benefits both teachers and students; most students are unaware, for instance, that intonation can accomplish all of these functions. For teachers, it can be very helpful to have these insights into learners' attitudes about intonation at the beginning of a course and can demonstrate that the task of teaching marked intonation is much larger than teachers may otherwise have anticipated. Since learners' underlying beliefs about intonation affect the strategies they will use when listening to English and making sense of speakers' intended meaning, learners should be directed toward an appreciation of the pragmatic functions of marked intonation. Teachers may be similarly unaware that students do not realize these facts. With this metacognitive framework in mind, however, teachers can focus students' attention on to the pragmatic functions of intonation in every exercise, and on every page of the textbook, throughout the semester. Including metacognitive assessments such as these at the beginning and end of instruction may also reveal real metacognitive progress that learners have made, though their pronunciation may not yet be approaching the target.

Use reading and inferencing to help scaffold learners' metacognitive understanding of intonation

L2 learners who have taken the TOEFL/IELTS or prep classes for these exams are familiar with the concept of inferencing from their exam preparation; they may not be familiar with the punctuation or aural signals that accompany the specific inferencing required when interpreting intonation (italics or marked intonation). Nevertheless, teachers can use the concept of inferencing to get learners used to

focusing on speaker intent, rather than more narrowly on surface level interpretations of just a speaker's words.

Vandergrift and Goh (2012) note that learners who have reached a threshold level of proficiency sufficient to segment words in connected speech still fall short in interpreting intended or implied meanings, reporting "understanding the words but not the message" (2012: 22). In this context, explicit instruction on intonation and its pragmatic effects can aid learners in listening courses as well as in general communicative contexts.

Add a metacognitive layer on to any pronunciation instructional materials

Teachers can bring this metacognitive approach to whatever set of textbooks or materials they are using. For every exercise or activity, teachers should be able to articulate the reasons why the particular intonation contour matters. Teachers then need to prompt students to articulate these reasons themselves.

Example:

"Look at these conversations. In some of the sentences, the focus word is circled. Decide which word you think would be the focus word in each of the other sentences. Circle it."

A: Is this 549-6098? [8 is circled]
B: No, this is 549-6078. (Hewings and Goldstein 1999: 107)

In this example from *Pronunciation Plus* (Hewings and Goldstein 1999) learners would have no trouble circling the "7" in the second line, as directed. The directions continue, prompting them to listen to the dialogue read aloud and "work in pairs and say the conversations together" (1999: 107). Ideally, in between marking up the sentence and reading it aloud, learners would be able to explain their choice, describe the way their pronunciation should signal the focus on the "7" and even reflect on the way their L1s would signal this change of focus. The exercise in itself is not problematic, but it assumes that teachers will direct learners' attention to how English intonation functions in this particular case.

Some books include this metacognitive aspect already, with direction lines that prompt learners to reflect on the functions of different intonation contours. For example, in *Clear Speech*, Judy Gilbert (2012) explicitly asks learners to "explain why the speaker emphasizes structure words in lines 3 [and] 5":

Example:

1. A: Do you think food in this country is ex**pen**sive?
2. B: **No**, not **rea**lly.
3. A: Well, **I** think it's expensive.
4. B: That's because you eat in **res**taurants.
5. A: Where do **you** eat?
6. B: At **home**. (2012: 73)

When discussing a sample sentence like this from a textbook, students would ideally be able to say that they notice extra stress and intonation on the underlined words and that fall–rise intonation on "I" in line 3 and "you" in line 5 implies a contrast between the two speakers. This oral recognition on students' part is the moment of learner "uptake" or, alternatively, "noticing" (Couper, Chapter 23 in this volume, citing Schmidt (2001)) the key suprasegmental feature at work in intonation instruction. The instruction does not end when learners are able to produce the marked intonation on demand, but when they are able to correctly interpret it in context. In short, if classroom materials do not prompt learners to move to the metacognitive level, teachers must.

Supplement books/real-world materials with specific examples that will focus learners on key metacognitive points

Along with listening/speaking or pronunciation textbooks, teachers may want to use authentic materials showing learners intonation "in action" – in videos, podcasts, and other contexts. These materials can be engaging and useful and can give learners many additional contexts for practice, but are best used in the service of learner metacognition, rather than as an end in and of themselves.

Specific examples (such as "The *teacher* didn't grade your exam," discussed above) that dramatically highlight the particular effects of intonation can serve as a supplement to these authentic materials. With examples like this, learners have the chance to see that intonation is so important in English that not attending to it can lead to interpretive errors on their part. While some authentic materials used in the classroom make intonation appear "decorative" (Gilbert 2014: 125), the fact that intonation can "undermine" (Wichmann 2005: 229) the words in an utterance can be revelatory to learners (and their teachers) and can suddenly prompt metacognitive realizations on their part.

Conclusion

In conclusion, we can see that textbooks and pedagogical materials on intonation have indeed improved, dramatically so, in the years since Levis (1999) noticed a divide between research and practice. However, there may still be practical problems in the day-to-day classroom implementation of these excellent materials. As we have argued, teachers can use textbooks that present intonation concepts clearly. They can use authentic materials and engaging, interactive activities. They can even get learners to produce the correct intonation contours on demand, but intonation instruction can still utterly fail if learners have not grasped the pragmatic importance of intonation for communication in English. Learners do not have to adopt the intonation contours characteristic of English into their daily speech, but they do need to be able to recognize these contours when they hear them, notice their role in signaling speaker intent, and discern the underlying meaning or

implication that they convey. Teachers play an important role in getting learners to this point, and intonation instruction should focus not just on grammar and emotion but also on implicature, and must go beyond the productive level to the metacognitive.

REFERENCES

Benz, C. and Dworak, K. 2000. *Tapestry Listening and Speaking 1*, Boston, MA: Heinle & Heinle.

Berko Gleason, J. and Ratner, N.B. 2009. *The Development of Language*, 7th edition, Boston, MA: Pearson Allyn & Bacon.

Beun, R.J. 1990. The recognition of Dutch declarative questions. *Journal of Pragmatics* 14: 39–56.

Bolinger, D. 1972. Accent is predictable (if you're a mind-reader). *Language* 48: 633–644.

Breitkreutz, J., Derwing, T.M., and Rossiter, M.J. 2001. Pronunciation teaching practices in Canada. *TESL Canada Journal* 19: 51–61.

Brown, S. 2011. *Listening Myths: Applying Second Language Research to Classroom Teaching*, Ann Arbor, MI: The University of Michigan Press.

Celce-Murcia, M., Brinton, D., Goodwin, J., and Griner, B. 2010. *Teaching Pronunciation: A Course Book and Reference Guide*, 2nd edition, New York: Cambridge University Press.

Chamot, A.U. 1995. Learning strategies and listening comprehension. In: *A Guide for the Teaching of Second Language Listening*, D. Mendelsohn and J. Rubin (eds.), 13–30, San Diego, CA: Dominie Press.

Collier, W. and Hubbard, T. 2001. Musical scales and evaluations of happiness and awkwardness: effects of pitch, direction, and scale mode. *American Journal of Psychology* 114(3): 355–375.

Couper-Kuhlen, E. 2001. Intonation and discourse: current views from within. In: *The Handbook of Discourse Analysis*, D. Schiffrin, D. Tannen, and H.E. Hamilton (eds.), 13–34, Oxford: Blackwell.

DeCasper, A.J. and Fifer, W.P. 1980. Of human bonding: newborns prefer their mother's voices. *Science* 208: 1174–1176.

DeCasper, A.J. and Spence, M.J. 1986. Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior and Development* 9: 133–150.

Foote, J.A., Holtby, A.K., and Derwing, T.M. 2011. Survey of the teaching of pronunciation in adult ESL programs in Canada, 2010. *TESL Canada Journal* 29: 1–22.

Gilbert, J.B. 1993. *Clear Speech: Pronunciation and Listening Comprehension in North American English*, 2nd edition, New York: Cambridge University Press.

Gilbert, J.B. 2012. *Clear Speech: Pronunciation and Listening Comprehension in North American English*, 4th edition, New York: Cambridge University Press.

Gilbert, J. 2014. Myth 4: intonation is hard to teach. In: *Pronunciation Myths: Applying Second Language Research to Classroom Teaching*, L. Grant (ed.), Ann Arbor, MI: University of Michigan Press.

Goh, C. 2008. Metacognitive instruction for second language listening development; theory, practice and research implications. *RELC Journal* 39(2): 188–213.

Grant, L. 2010. *Well Said: Pronunciation for Clear Communication*, 2nd edition, Boston: Heinle & Heinle.

Grant, L. 2014. Prologue to the myths: what teachers need to know. In: *Pronunciation Myths: Applying Second Language Research to Classroom Teaching*, Ann Arbor, MI: The University of Michigan Press.

Gumperz, J. 1982. *Discourse Strategies*, Cambridge: Cambridge University Press.

Hagen, S. 2000. *Sound Advice: A Basis for Listening*, 2nd edition, White Plains, NY: Pearson.

Hahn, L.D. 2004. Primary stress and intelligibility: research to motivate the teaching of suprasegmentals. *TESOL Quarterly* 38(2): 201–223.

Hahn, L.D. and Dickerson, W.B. 1999. *Speech Craft: Discourse Pronunciation for Advanced Learners*, Ann Arbor, MI: The University of Michigan Press.

Hewings, M. and Goldstein, S. 1999. *Pronunciation Plus: Practice Through Interaction*, New York: Cambridge University Press.

Hirst, D. and Di Cristo, A. 1998. A survey of intonation systems. In: *Intonation Systems: A Survey of Twenty Languages*, D. Hirst and A. Di Cristo (eds.), 1–44, Cambridge: Cambridge University Press.

Jenkins, J. 2000. *The Phonology of English as an International Language*, Oxford: Oxford University Press.

Jenkins, J. 2002. A sociolinguistically-based, empirically-researched pronunciation syllabus for English as an International Language. *Applied Linguistics* 23(1): 83–103.

Karmiloff, K. and Karmiloff-Smith, A. 2001. *Pathways to Language: From Fetus to Adolescent*, Cambridge, MA: Harvard University Press.

Levis, J. 1999. Intonation in theory and practice, revisited. *TESOL Quarterly* 33(1): 37–63.

Levis, J. 2005. Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly* 39(3): 369–377.

Lieberman, P. 1967. *Intonation, Perception, and Language*, Cambridge, MA: MIT Research Monograph 38.

Lyster, R. and Ranta, L. 1997. Corrective feedback and learner uptake: negotiation of form in communicative classrooms. *Studies in Second Language Acquisition* 20: 37–66.

Mendelsohn, D. 1998. Teaching listening. *Annual Review of Applied Linguistics* 18: 81–101.

Mennen, I., Schaeffler, F., and Doherty, G. 2012. Cross-language difference in f0 range: a comparative study of English and German. *Journal of the Acoustical Society of America* 131(3): 2249–2260.

Miller, S.F. 2006. *Targeting Pronunciation: Communicating Clearly in English*, 2nd edition, New York: Houghton Mifflin.

Molinsky, S.J. and Bliss, B. 2002. *Side by Side*, White Plains, NY: Pearson Longman.

Moon, C., Cooper, R., and Fifer, W. (1993). Two-day-olds prefer their native language. *Infant Behavior and Development*, 16, 495-500.

Paunović, T. and Savić, M. 2008. Discourse Intonation – making it work. In: *As You Write It: Issues in Literature, Language, and Translation in the Context of Europe in the 21st Century*, S. Komar and U. Mozetić (eds.), vol.1–2: 57–75.

Pennington, M.C. and Ellis, N.C. 2000. Cantonese speakers' memory for English sentences with prosodic cues. *The Modern Language Journal* 84(3): 372–389.

Pickering, L. 2012. Intonation. In: *The Routledge Linguistics Encyclopedia*, 3rd edition, K. Malmkjaer (ed.), 280–286.

Reed, M. and Michaud, C. 2005. *Sound Concepts: An Integrated Pronunciation Course*, New York: McGraw-Hill.

Rost, M. 2005. L2 listening. In: *Handbook of Research in Second Language Teaching and Learning*, E. Hinkel (ed.), 503–527, Mahwah, NJ: Erlbaum.

Schmidt, R.W. 2001. Attention. In: *Cognition and Second Language Instruction*, P. Robinson (ed.), Cambridge: Cambridge University Press.

Searle, J. 1969. *Speech Acts: An Essay in the Philosophy of Language*, Cambridge: Cambridge University Press.

Searle, J. and Vanderveken, D. 1985. *Foundations of Illocutionary Logic*, Cambridge: Cambridge University Press.

Spence, M. and DeCasper, A. 1987. Prenatal experience with low-frequency maternal-voice sounds influence neonatal perception of maternal voice samples. *Infant Behavior and Development* 10: 133–140.

Tomlinson, J.M. and Bott, L. 2013. How intonation constrains pragmatic inference. In: *Proceedings of the 35th Annual Conference of the Cognitive Science Society*, M. Knauff, M. Pauen, N. Sebanz, and I. Wachsmuth (eds.), 3569–3574, Austin, TX: Cognitive Science Society.

Uldall, E. 1964. Dimensions of meaning in intonation. In: *In Honor of Daniel Jones*, D. Abercrombie, D.B. Gry, P.A.D. MacCarthy, N.C. Scott, and J.L.M. Trim (eds.), 2712–279, London: Longmans.

Vandergrift, L. and Goh, C. 2012. *Teaching and Learning Second Language Listening: Metacognition in Action*, New York: Routledge.

Wells, J.C. 2006. *English Intonation: An Introduction*, Cambridge: Cambridge University Press.

Wharton, T. 2012. Prosody and meaning: theory and practice. In: *Pragmatics, Prosody and English Language Teaching*, J. Romero-Trillo (ed.), 97–117, Springer.

Wichmann, A. 2005. Please – from courtesy to appeal: the role of intonation in the expression of attitudinal meaning. *English Language and Linguistics* 9(2): 229–253.

26 Integrating Pronunciation into the Language Classroom

LAURA SICOLA AND ISABELLE DARCY

Introduction

Few language students are fortunate enough to have a class that is dedicated to the sole focus of studying pronunciation, and even fewer are able to take such a class with a teacher who is genuinely knowledgeable about English phonology and pronunciation pedagogy. Pronunciation is frequently relegated to the occasional side lesson in the context of a broader oral communication course or omitted entirely from the curriculum. This unsystematic and infrequent approach to pronunciation is insufficient for many learners to orally convey their messages intelligibly and effectively. This chapter will begin by looking at challenges faced by many "regular" ESL/EFL teachers regarding teaching pronunciation to frame the subsequent suggestions made for making pedagogical connections between pronunciation and teaching the other skill areas (speaking, listening, reading, and writing) within a communicative framework. To further contextualize the suggested strategies, we review the theoretical underpinnings of the value of communicative tasks in pronunciation instruction in an effort to guide practitioners in making pronunciation targets an inherent part of every lesson.[1]

Challenges

Pronunciation difficulties in a second language (L2) can seriously impede intelligibility. Developing fluent speech and intelligible pronunciation plays a crucial role for L2 learners' social and economic integration, such as for L2 learners of English who live in an English-speaking environment. Lack of intelligible pronunciation is also accompanied by comprehension difficulties when L2 learners listen to spoken English.

The Handbook of English Pronunciation, First Edition. Edited by Marnie Reed and John M. Levis.

At the same time, the prospect of resolving pronunciation difficulties in the classroom presents a considerable challenge. This is mainly due to two reasons: (1) intelligible pronunciation is difficult to learn for most adults and (2) intelligible pronunciation is difficult to teach due to a lack of teacher preparation, limited availability of materials, and ineffective teaching. By "ineffective," we refer to either a heavily form-focused instruction (e.g., minimal pair drills) or an exclusively meaning-focused approach without explicit attention to phonological form.

The former, identified as a lack of contextualization of pronunciation instruction (Bowen 1972), is characterized by the exclusion of meaning integration, with no or little carryover from the classroom lesson to any external/spontaneous conversation and no integration of pronunciation targets into spontaneous speech. At the opposite extreme is an exclusive focus on meaning, favored by typical communicative language teaching methods. Exclusively meaning-focused instruction offers too few opportunities for repetition of familiar materials because of the primary allocation of attention to higher levels of information exchange (Segalowitz and Hulstijn 2005). Thus, sole reliance on this approach fails to foster automatization of phonological and phonetic processing in the L2.

In this chapter, we argue that making pronunciation targets an inherent part of every lesson could represent an effective solution to carryover and automaticity issues. However, there are three major challenges to integrating pronunciation teaching into the broader language classroom.

The first challenge is the lack of teacher training in pronunciation. Many teachers do not feel confident in their knowledge about pronunciation or in their ability to teach it (Foote, Holtby, and Derwing 2011). Often this is because they received minimal practical training in this area, if any, as many TESOL training programs incorporate little to no pedagogical training around pronunciation. Courses often offer a brief formal introduction to phonology on a theoretical level, but practical application is usually limited to activities such as transcribing recorded speech samples using the IPA. While this may provide teachers with a deeper understanding of English pronunciation, it does not provide them with an understanding of how to teach it. A related challenge is that, particularly in many outer/expanding circle countries (Kachru 2005), a number of teachers who are non-native speakers of English (NNSs) lack confidence in their ability to successfully model English pronunciation, perhaps feeling that their own pronunciation is too deviant from a "target-like" pronunciation. Non-native teachers in many countries often teach more "metalanguage-heavy" classes, i.e., teaching *about* English (grammar rules, vocabulary lists, etc.) but *through* their native language, thus providing very few opportunities for students to hear their teachers modeling spoken English. As a result, the cycle becomes self-perpetuating. Students who have only experienced L2 learning in an educational system that prioritizes the passing of standardized tests, and in a classroom context that is teacher-centered and primarily conducted in the L1 with little opportunity to hear or practice English pronunciation, later become English teachers in the same system and are likely to use similar teaching methods, all without acquiring and therefore using skills related to pronunciation instruction.

While a highly intelligible NNS teacher is an appropriate model, there are also many means by which NNS teachers can expose their students to various native-speaker models, by drawing upon an expanding number of resources, recordings, and other audio or audiovisual materials spoken by native speakers. Phonological differences between the NNS teacher's speech and that of native speakers clearly do not preclude the teacher from providing explanations and feedback, but many lessons may also contain recordings from native speaker utterances. It is beneficial to expose students to as high a number as possible of different native speakers' voices, so that their perceptual learning and listening skills become more robust (e.g., Bradlow et al. 1999).

The second challenge is that pronunciation is rarely assessed systematically in proficiency placement tests, whether in a community language program, a university-level intensive English program, or in a general primary or secondary school setting.[2] A problem is that pronunciation is difficult and time-consuming to assess objectively, and standardized tools are not yet available. It cannot be done via simple multiple-choice means and generally requires the audio-recording of a speech sample for later evaluation or for individual interviews to be conducted and assessed in real-time. Evaluators often do not have the phonological training to evaluate the samples and identify what targets to prioritize for particular students.

This holds an important implication in terms of integrating pronunciation into other language classes. No matter how much a program tries to group students by proficiency levels, overall proficiency or syntactic accuracy is not clearly correlated to phonological accuracy. Thus, students of varying pronunciation levels will be in the same classes, requiring teaching and assessment of pronunciation to be somewhat individualized. If this is the case, finding pronunciation lessons where the target form is selected to fit a well-defined "proficiency level" may be moot, since at any given moment, the students may need assistance producing a phonological target form that is inherently relevant to whatever other language forms and skills are being incorporated in the day's lesson. Darcy, Ewert, and Lidster (2012) did, however, outline areas of phonological targets that would be more appropriate to address with students of different proficiency levels, which can help the teacher prioritize the elements that are both developmentally suitable for different students and relevant to the other lesson objectives of the day.

The third challenge is related to a late introduction of specific pronunciation instruction, perhaps due to its perceived need for metalinguistic description, which requires specialized vocabulary and which may seem too advanced for beginning students to handle. A tendency common to many programs is therefore to make pronunciation an elective or an "advanced" class, instead of introducing pronunciation components in the early levels. We argue that it is essential for pronunciation to be introduced early, frequently, and as a regular component – large or small – of every lesson, avoiding metalinguistic or technical language in the early proficiency levels. Helping students perceive and produce more target-like pronunciation patterns from the start appears more effective than keeping students reinforcing non-target-like pronunciation over years, which then needs to be unlearned under greater effort. As Darcy, Ewert, and Lidster (2012) delineate,

pronunciation as an instructional focus should be "embedded, both within the curriculum as a whole, and within each lesson locally: pronunciation is not taught separately from, but rather becomes an integral part of, general language instruction" (2012: 95). Our challenge, then, is to help practitioners identify ways to execute this call to action. To that end, we now look at ways in which teachers at nearly any level and in any context can incorporate explicit attention to phonological forms, both proactively and incidentally (Ellis, Loewen, and Basturkmen 2006) in the context of other language lessons throughout the day.

Before specifically discussing strategies and techniques to incorporate pronunciation targets into other areas, we review a communicative framework for teaching pronunciation (Celce-Murcia et al. 2010) as a potentially useful framework from which to draw specific pedagogical elements.

Form-focused communicative language teaching

One central component of developing fluency and accuracy in pronouncing the L2 is automaticity of phonological and phonetic processing. According to Segalowitz and Hulstijn (2005), typical methods that provide the repetition necessary for automaticity to develop fail to promote learning because of the highly decontextualized nature of the repeated materials (2005: 383); at the same time, exclusively meaning-oriented activities fail to provide the repetition necessary for automatization. Gatbonton and Segalowitz (1988: 478) suggest that it is possible to promote (phonological) acquisition through activities requiring a dual focus on both form and meaning, i.e., activities that are inherently repetitive yet genuinely communicative (see also Canale and Swain 1980). With practice, attention to form becomes automatized (Gatbonton and Segalowitz 1988; Trofimovitch and Gatbonton 2006). Applied to pronunciation, to ensure that attention to form is indeed maintained as learners focus more on meaning, there will ideally be a design feature that requires accurate perception and/or production of the target form as essential to the successful completion of the activity (Loschky and Bley-Vroman 1993).

The communicative framework for teaching pronunciation outlined by Celce-Murcia et al. (2010) offers a way to achieve such an integration of form and meaning. It aligns pronunciation classroom practices with the tenets of communicative language teaching, in gradually shifting the scope of the focus of attention over the course of the work on a given topic. The framework defines five phases: 1. Description and analysis, 2. Listening discrimination, 3. Controlled practice, 4. Guided practice, 5. Communicative practice (Celce-Murcia et al. 2010: 44–49). Starting with a detailed focus on metalinguistic description and analysis, attention is gradually shifted towards incorporating more meaning, while retaining focus on the form. This is mainly achieved through a sequence of activities in which meaning becomes gradually more important, and for which corrective feedback is planned accordingly (Saito and Lyster 2012; Reed 2012).

One way to increase the likelihood that students fully engage in attending to both form and meaning is through the use of interactive tasks. "Tasks", as a subset

of the more general "activities", have been defined in various ways. For our purposes, we draw from the work of Willis and Willis (2007) and Pica, Kanagy, and Falodun (1993). Willis and Willis define a task as an activity that (a) engages learners' interest, (b) has meaning as a primary focus rather than form, (c) requires completion, (d) has a specific outcome on which "success" is based, and (e) relates to the "real-world" (1993: 13). Pica, Kanagy, and Falodun's (1993) typology of tasks looks more narrowly at the features of a task that are most likely to maximize negotiated interaction between learners. This is achieved when a task requires participants to request and provide uniquely held information, seek clarification regarding L2 input that they do not understand, and modify their utterance when they receive similar clarification requests in response to their own interlanguage production, all with the aim of reaching a mutually understood and accepted communication goal. Accordingly, the overarching function of a genuinely communicative task is to have students engage in work that is authentic in its relationship to real-life events, the outcome of which is independent of the use of language for its own sake.

However, we see two main ways in which this framework can be applied to incorporate pronunciation targets into any language lesson. The teacher can either proactively select pronunciation targets around which to organize a lesson (see Sicola 2009 for a discussion of proactive selection of phonological target forms in the context of interactive communicative tasks) or he or she can systematically address pronunciation issues as they arise in students' authentic production while completing a task. Pronunciation becomes integrated when the successful task completion crucially *depends* on target-form accuracy.

Part of the challenge in proactively teaching pronunciation forms for students is that it is difficult to create authentically communicative, interactive activities, in which accuracy of pronunciation-related target forms (segmental or suprasegmental) is essential to successful task completion (Loschky and Bley-Vroman 1993). Sicola (2009) gives an overview of this challenge and demonstrates how her example of a map task combines the meaning-focused quality of communicative tasks with pre-selected phonological targets in a way that will produce target-form-related negotiated interaction among student participants so that target-form accuracy becomes essential to successful task completion. Within the communicative framework, Celce-Murcia and colleagues (2010) characterize such activities as "guided practice".

While this may be an ideal situation, such tasks are typically not readily available to most teachers and can be time consuming to design. However, there are ways to compensate for this gap and adjust task conditions so that negotiated interaction and attention to phonological target forms are promoted in the context of broader language tasks. Willis and Willis (2007) provide an extensive list of task types to be used with students of varying proficiency. These tasks do not need to be complex; they can be as straightforward as brainstorming, guessing games, memory challenges, sequencing, ranking, classifying, creating timelines and tables, etc. For more advanced levels, more complex types include problem-solving tasks, comparison and contrast analyses, creative story-telling, and projects. Their "task

generator" (2007: 108) offers a useful framework for incorporating any of seven categories of task types into language lessons on any particular topic or target form, all of which can be modified to meet the needs of specific proficiency levels. Importantly for our purposes, pronunciation targets can be woven into these tasks from the beginning. For example, when creating a timeline, ordinal numbers are a natural and useful construct (Loschky and Bley-Vroman 1993), and words such as *first*, *second*, *third*, *fourth*, etc., all include complex coda structures, which can be difficult for learners to pronounce, and thus attention to this issue can be included and reinforced throughout the task and lesson overall. By requiring the repetition of target forms in a variety of genuinely communicative, applied contexts, these tasks correspond to the ideal balance for automatization and carryover outlined by Gatbonton and Segalowitz (1988).

Pica, Kanagy, and Falodun's typology of tasks (1993) further describes the likelihood for different task types of maximizing the participants' negotiated interaction, and how to adjust task conditions in order to increase this likelihood, whether jigsaw tasks, opinion exchange tasks, or decision-making tasks. To the extent that language teachers can incorporate these tasks and criteria into their lessons, there is a much greater likelihood that learners will produce ample authentic language. As these tasks are intended to serve a greater communicative function, and are not typically pronunciation focused, teachers should be able to incorporate them into their lessons by helping students deliberately work to produce the target form more accurately as a step toward acquisition.

Using the communicative framework to integrate a pronunciation component into other lessons

In this section, we look at some ways in which pronunciation can be integrated with other areas, especially vocabulary, spelling, grammar, listening/speaking, reading, and writing. The suggestions are meant to be illustrative, not exhaustive.

Vocabulary

One area in which teachers working with any level, content, skill or population all share an excellent opportunity and obligation to address pronunciation is with the introduction of new vocabulary. We offer suggestions for helping students meet the challenge of learning target-like pronunciation of the lexicon.

One of the first things students focus on when learning new vocabulary is how the words are spelled. This gives the teacher an opportunity to address patterns of pronunciation and orthography (Celce-Murcia et al. 2010). The influence of spelling on *literacy* skills (word recognition, vocabulary learning, writing) is well known, but its influence on the emergent *sound system* is also important and should not be overlooked (Prator 1971; Escudero, Hayes-Harb, and Mitterer 2008). Spelling is often considered ancillary to other goals pertaining to vocabulary, syntax, or fluency development, yet addressing orthography can be a very important part of

developing intelligible pronunciation. While there are often exceptions, many simple patterns can be productively taught. More importantly, it is essential for teachers to recognize that some learners will attempt to make sense of the system whether they have help or not. Therefore, providing guidance and awareness will help students who make erroneous connections between graphemes and phonemes. Indeed, since learners have a disadvantage in inferring patterns because they lack native-like phonological awareness, mis-mapping is likely. By making it clear from the beginning that few rules apply without exceptions, confusion is not the most likely outcome. Gilbert (2001) and Celce-Murcia et al. (2010) provide many helpful suggestions for teaching connections between spelling and pronunciation. (For a thorough overview of the relationship between orthography and pronunciation, see Markham 1997 and Dickerson's Chapter 27 in this volume for strategies in explicitly using orthography to teach pronunciation.)

One example activity for beginning students is focusing on the different sounds associated with the letter <c> (/s/, /k/) in words such as *city* and *cat* respectively. The predictive rule is straightforward, with <c> pronounced as [s] before the letters <e>, <i> and <y> and [k] everywhere else. Of course, the same letter <c> when combined with other letters such as <h> is typically associated with a new sound /tʃ/ in words such as *child*. Like anything in English, there are exceptions to the pattern, but the regularity will help leaners connect what they see and what they say. Patterns can be addressed using words known to students and having them form categories first, before adding the new vocabulary into these categories. Using pairs or groups encourages students to make their hypotheses explicit and gives opportunities for corrective feedback and/or praise. Applying it to unknown words (such as vocabulary in subsequent readings) can help convince students of the usefulness of the activity.

Another example of a "learner-driven mis-mapping" and a useful pattern to learn is the pronunciation of <ay> and <ai>, which almost always represent [eɪ], but are typically misconstrued as [aj]. There are few exceptions to the rule in American English, such as the third person form *says* [sɛz]; the rest are mostly rare or unassimilated loanwords. Once students understand this pattern, they can use it to make the connection between words they know well, such as *today*, and new words they encounter, such as *allay* (which might initially be read – and even understood – as *ally* [ˈælaj]), thereby improving students' independent ability to accurately predict and produce target-like pronunciation of new words.

From spelling, syllabification and stress patterns are a logical next step. Attention to lexical stress patterns, at least, should be an inherent part of the introduction of new polysyllabic words. Other factors related to lexical stress also play an important role in intelligibility (Benrabah 1997; Derwing, Munro, and Wiebe 1998; Field 2005; McCrocklin 2012; see also Derwing and Munro, Chapter 21, and Cutler, Chapter 6, in this volume), arguably because stress placement has a direct effect on phonetic production, such as in the words *democrat* (/ˈdɛ mə ˌkɹæt/) and *democracy* (/də ˈmɑ kɹə ˌsi:/). Because stress is largely a redundant feature in English, English listeners may perceive stress not only through length or pitch but also through segmental production.

Similar strategies can be applied to phrase-level stress, given that inappropriate stress assignment, either at the word or phrase level, may result in an unintelligible production (Derwing and Rossiter 2002; Zielinski 2008). Students might benefit from discovering that patterns they know from words (e.g., *today*) can also be applied to phrases such as *at work* or *at home.* This approach can then be extended to frequent phrases and syntactic chunks that have fixed stress patterns, such as *I wish I'd known* (Field 2014). Although explaining the meaning of words and phrases is important, such enriched vocabulary lessons that increase phonological awareness can help students to remember the words and phrases accurately. One possible way to implement this suggestion is to use the notation system proposed by Murphy and Kandil (2004) to identify stress placement in new words. Their system uses number sequences such as 3-2 to indicate the accentuation patterns. Words like *assessment* or *specific* have the same 3-2 pattern, where the initial number indicates the number of syllables and the second number indicates which syllable carries the primary stress.

These suggestions can be extended to content–area classes, whether Language for Specific Purposes (LSP) or a non-language focused class, such as a high school science class. Any of the above connections to pronunciation are still valid and should be incorporated when helping NNSs develop their overall academic or professional language proficiency. By nature, the classroom is a place in which the learning of subject-specific vocabulary and discourse styles is an expected result. To the extent that teachers are aware of the parts of speech represented, they could explicitly point out challenging segments and lexical- and phrasal-stress patterns in new vocabulary words and collocations.

For example, they can explain how suffixes influence stress placement using the example of words ending in *-ology*, which always receive primary stress on the first syllable of the suffix itself, as in *biology.* This is something most teachers can learn and they should hold the students and themselves accountable for producing the word with well-placed stress. This can be reinforced during any oral activities, ranging from times when students are reading aloud from a textbook or from their own compositions to open classroom discussion or more formal oral reports and presentations.

Grammar

There is unquestionably a link between some grammatical structures and pronunciation (Celce-Murcia et al. 2010). In surveying a range of student ESL textbooks, most of which were either speaking/listening-focused or integrated all four skill areas, we noted that several had explicit activities, instructions, or footnotes pertaining to the relationship between pronunciation and grammar. Most often this occurred in the context of introducing a new grammatical construct (or within the first few exercises), if the successful use of the target form was at least partly dependent upon its phonetic realization. This link should be made upon introducing these forms and reinforced whenever possible, once a form has been taught. For example, typically, the regular noun and verb endings *-s* and *-ed* are

inaccurately realized in spontaneous speech or in reading aloud, either by omission (e.g., *places* being pronounced /pleɪs/) or erroneous addition of an extra syllable (e.g., *baked* being pronounced /ˈbeɪkəd/), along with errors in voicing of the final consonant. This may be due to L1 coda or coda cluster syllable structure constraints, as well as lack of integrated knowledge of the allomorphic rules (Lardiere 2003; Jiang 2007). Explicit instruction about the rules governing allomorphs that provides both sufficient opportunities for production of the target structure and corrective, form-focused feedback has been shown to help learners convert their explicit knowledge of rule-governed structures to spontaneous production (Reed 2012; Yang and Lyster 2010).

Discussion of suffixes, whether inflectional or derivational morphemes, is a natural connection to parts of speech. At more advanced levels, helping students recognize the relationship between pauses, phrase-level stress, parts of speech, and thought groups can lead to significant improvements in intelligibility (van Loon 2002). Feasibly, this relationship can be introduced to students at a lower proficiency level than van Loon's students, in a less metalinguistic way. By modeling and recasting simple patterns that incorporate the targeted grammatical form, the teacher can draw students' attention to rhythm, pausing, and stress patterns during oral practice of the activity.

Table 26.1 outlines some grammatical constructs in English whose successful oral application requires accurate, rule-based phonetic realization.

Speaking and listening

Pronunciation should play a central role in the development of oral skills, regardless of the specific focus of an activity (vocabulary, grammar practice, etc.). There is an inextricable link between speaking and listening: they are linked interactively, as by nature oral activities require the message to be pronounced intelligibly and perceived accurately if they are to be completed successfully, and they are linked internally, as speaking and listening can serve as an auditory feedback loop, with a student's speech serving as his or her own input (Reed and Michaud 2011).

The relationship between speaking and listening can also be viewed as a mutually beneficial one in terms of acquisition: there is substantial evidence that improved perceptual/listening abilities can transfer to production/speaking (Rvachew, Nowak, and Cloutier 2004). For example, studies using high-variability training paradigms have generally shown that in controlled laboratory conditions perceptual training can cause L2 learners to improve not only their perception but also, critically, their production of segmentals (e.g., Bradlow et al. 1999) and even suprasegmentals (e.g., Wang, Jongman, and Sereno 2003). Conversely, pronunciation practice can also help developing listening comprehension skills, as suggested by Gilbert (1995). Specific empirical evidence is limited, but it appears that learning to correctly realize word stress, vowel reduction, and word linking patterns might help students segment fluent speech and recognize words more accurately in native speakers' utterances (Diane Poisson, personal communication, November 6, 2011).

Table 26.1 Grammatical forms with direct connections to pronunciation.

Grammatical form	*Pronunciation targets*	*Pronunciation-related example*
Past tense/past participle *-ed* Plural *-s* Third person singular *-s*	Allomorphic variation; simple and complex codas; extra syllable	walked /wɔkt/, agreed /əgɹi:d/, wanted /wɑntəd/ cats /kæts/, dogs /dɔgz/, horses /hoɹsəz/ eats /i:ts/, seems /si:mz/, smashes /smæʃəz/
Contractions Content words versus function words	Reductions; simple and complex codas	does not /dəz ˈnɑt/ → doesn't /ˈdʌznt/ I'll talk <u>to you</u> later. → /tə jə/
Nouns and verbs Compound nouns and descriptive phrases Compound nouns and phrasal verbs Suffix patterns and parts of speech	Lexical stress	contract (n.) /ˈkɑ:ntrækt/ vs. contract (v.) /kənˈtrækt/ (the) ˈ*White* House (where the US president lives) versus (any generic) white ˈ*house* ˈ*print*-out (cn) versus print ˈ*out* (something) (pv) Words ending with *-ic* (adj.) always stress the penultimate syllable, i.e., the syllable prior to *–ic*. E.g., demoˈcratic, symˈbolic, euˈphoric, systeˈmatic
Questions: giving someone a number of choices Asking for new information versus clarification Tag questions: inquiry versus accusation	Intonation	Do you want cake/ or ice cream\? (choose one of the two) Do you want cake/ or ice cream/? (either, neither, or both) What time is the meeting\? (first inquiry) What time is the meeting/? (clarification or confirmation) You're not coming to the party\, are you/? (inquiry) You're not coming to the party\, are you\? (accusation)
Correction and emphasis	Phrasal stress	– "I think the plane lands at three *o'clock*." – "No, I heard it lands at *four* o'clock."
Clauses, thought groups	Pausing, rhythm, intonation	My *daughter* <u>who likes to *swim*</u> is having a *POOL* party.\ (necessary information, as opposed to another daughter who does not like to swim) My *daughter*, (.) <u>who likes to *swim*</u>,(.) is having a <u>*POOL*</u> party.\ (useful but unnecessary information about the daughter)

Listening should be viewed as an interactive and interpretive process, rather than a passive one (van Loon 2002; Murphy 1991). Empirical evidence shows that active listening tasks that direct students' attention to noticing more nuanced details of pronunciation can be more effective than only oral practice activities in helping students to develop more target-like pronunciation (Counselman 2010; Pennington and Ellis 2000). The possible benefits of this practice may be further magnified by engaging learners in directed listening to their own speech.

Recording technology can be useful to facilitate such active listening activities (see Hincks' Chapter 28 in this volume). Even if students do not have access to a computer laboratory equipped with high-performance technology or advanced speech analysis software, most students are able to record and listen to their own speech with portable devices (e.g., smartphones or iPods). This enables them to listen to and analyze their speech more objectively rather than trusting their memory of what they said or how they said it. The teacher can then fruitfully draw their attention to certain target forms and features as they actively listen for areas of successful improvement and collaboratively set new goals for learning.

Reading and writing

One of the classroom practices with which many students have a "love–hate" relationship is reading aloud. On the one hand, all of the required text is already present in a target-like form, so there is less risk of making a grammatical or lexical error. Without needing to allocate cognitive resources to those issues, students are more able to attend to their pronunciation (Robinson 2001). On the other hand, reading aloud puts additional pressure on students to use more accurate pronunciation in front of the rest of the class. As previously mentioned, the complex relationship between orthography and pronunciation can both promote and inhibit target-like production while reading aloud. For example, the written symbols may remind students to produce sounds that they might otherwise forget when speaking freely, e.g., seeing the digraph <th> may remind them to produce the sound /θ/ (as in *think*) or /ð/ (as in *they*). Conversely, it is likely that the irregular spelling patterns of English will mislead students to mispronounce even known words (Levis and Barriuso 2012; Sicola 2009).

Because reading aloud is one of the most common activities experienced in the classroom, it offers a frequent and consistent opportunity for the teacher to draw students' attention to pronunciation. For instance, the teacher can include a reminder as part of the instructions or have students silently pre-read a passage in order to scan for and underline any words that include particular letters and combinations (e.g., <c>, <ough>, <ic>) and noun and verb endings, such as the plural *-s* or past tense *-ed*. (The fact that students often fail to articulate these endings even when reading from the printed word right on the page is evidence that their failure to use these forms correctly in conversation may be a phonological issue rather than a grammatical one.) Slightly more complex would be having students use their metalinguistic knowledge of which words or parts of speech carry most meaning and how to identify clause boundaries and thought groups in order to

mark the text accordingly in preparation (van Loon 2002; Reed and Michaud 2005). This pre-reading can also serve to address many other pronunciation topics and should become a routine in the classroom.

Since learners initially have difficulties processing form and meaning at the same time (Doughty and Williams 1998), such a pre-reading opportunity can also have a more multifaceted positive effect in that it will help learners process the meaning of the passage prior to reading it aloud. Once they have gained some familiarity with words and broader meaning, they are more likely to be able to include more target-like pronunciation, and the marks serve as visual cues to remind students to attend to segmental and/or suprasegmental details of those targets when they encounter them during the subsequent read-aloud. This progressive training in attending to both meaning and form at once is also likely to trigger more carryover and potentially narrow the gap between the learner's pronunciation patterns during "formal" read-aloud and "informal" free-speech activities (Archibald 1998; Major 1987; Segalowitz and Hulstijn 2005).

Teaching writing also provides opportunities to teach and practice a wide range of pronunciation targets, which can and should be incorporated regularly. For this purpose, we can group writing efforts into two broad categories, which we will refer to as "mechanical" writing skills and "discourse" writing skills. Mechanical writing skills include learning to form L2 symbols, i.e., letters and characters, if the L2 script is different from the L1 script; spelling and word construction; and sentence-level writing, typically to practise particular syntactic structures or vocabulary items. Discourse writing skills are at the composition level, creating paragraphs and beyond, and putting one's own thoughts into more extensive L2 text.

Starting with mechanical skills, particularly when working with beginners, basic assignments such as "write each character/word five times" are common. Pronunciation can be incorporated by having students name the symbols or the sounds they represent while the teacher monitors the assignment in real-time in the classroom. Even when reviewing basic assignments by comparing their work with a partner's, they can read their own or each other's work aloud and ultimately come to an agreement on whether or not the symbols, words, or answers are correct, a task characteristic that also maximizes negotiated interaction and provides more opportunities for attention to form (Pica, Kanagy, and Falodun 1993), including phonological targets.

Discourse writing skills frequently require engaging in some or all of the stages of the writing process, which may include tasks such as prompt-deconstruction, brainstorming, organizing and outlining, drafting, sharing and read-aloud, peer review, revision, editing, and finally publication (Williams 2003). The majority of these stages include opportunities for collaboration, thereby shifting the mode of learning from singularly text-based to oral exchange, enabling the incorporation of a variety of pronunciation targets. For example, interpreting a writing prompt collaboratively requires students to deliberate and reach an agreement, a good opportunity to practise suprasegmental strategies for clarification requests and making contrasts. Brainstorming gains momentum when done in groups and is a perfect context for list-intonation patterns, for example; similarly, narrative storytelling

as a pre-writing stage also requires attention to suprasegmental features that delineate thought groups (Levis and Grant 2003). Next, organizing the subsequent outline by deciding what brainstorm items to include and how to place them in the outline once again requires students to negotiate in order to reach an agreement. Debate would require contradiction and contrast, which become clearer and more powerful when spoken with correct intonation patterns. Sharing drafts can then be done once again by reading the compositions aloud to peers, rather than exchanging papers and reading them silently. Thus, at the very least, in the course of these discussions, the teacher can address pronunciation issues as they occur or can proactively weave in a deliberate focus on a relevant pronunciation target form.

Pronunciation in other content-area lessons

An increasingly common scenario in the United States and many other countries is the situation of younger immigrant, exchange, and otherwise international students of varying English proficiency levels enrolling in PreK-12 public schools, vocational/trade schools, and other educational programs in which there is often little or no formal ESL instruction or faculty trained in L2 pedagogy. Elementary school teachers who have self-contained classes and teach all subjects to their students and secondary or tertiary teachers of mathematics, science, art, economics, and other content areas are often the students' only source of formal guidance for language development. Academic literacy in any subject includes not only content knowledge but also the ability to *intelligibly communicate* one's understanding of that content. Teachers of all subjects need to recognize their agency in students' subject-specific language development, as well as the importance that target-like pronunciation plays therein. In considering this responsibility and how it would ultimately influence their teaching, we hope teachers will consider the various strategies and rationales we have offered, such as those pertaining to vocabulary development, for example, and find ways to incorporate them regularly into their lesson plans and student achievement expectations and outcomes.

Conclusion

Pronunciation is a very important component of oral communication and just like the other components of language it should be taught as part of an integrated, interdependent system. Pronunciation skills are interconnected with other areas such as listening comprehension, reading and writing, and grammar. Given these interconnections, it is crucial to address the pronunciation needs of students at an early stage and throughout the curriculum. In fact, improved pronunciation may help – and, conversely, persistently non-target-like pronunciation may interfere with – students' performance in all other areas of the curriculum.

It is our hope that this chapter will encourage practitioners and program administrators to recognize that pronunciation needs of students are best addressed across all curriculum areas; ideally, students' ability to recognize the relevance of pronunciation

across contexts is essential for their optimal success as L2 users. We hope to have encouraged practitioners to consider pronunciation as an integral part of L2 learning by having demonstrated that it is feasible to weave pronunciation targets into every lesson regardless of skill area. Preferred activities ideally combine a communicative purpose while promoting automaticity of phonological processing, a combination that is likely to enhance the effectiveness of pronunciation instruction.

Acknowledgments

The authors wish to acknowledge Ziwei Zhou for reviewing pedagogical resources to identify whether and how pronunciation was addressed in the context of other ESL textbooks, and to thank Vance Schaefer and Ryan Lidster for helpful feedback on previous drafts of this chapter.

NOTES

1 We understand that different teaching contexts will make it more difficult for some teachers to implement certain suggestions relative to others. Our goal is to demonstrate possibilities for connecting theory and practice in the hope that practitioners will adapt these ideas and examples in ways that can be best implemented in their own classrooms.

2 Of note, there are a number of university programs that do assess *global* speaking and listening skills upon enrollment (e.g., Michigan State University, University of Michigan, University of Iowa, Indiana University) and some also include a specific pronunciation rubric in their diagnostic assessments for incoming international freshmen, such as the Indiana English Proficiency Exam (Indiana University). Some Intensive English Programs also include specific pronunciation assessments (e.g., University of Iowa, Indiana University). For K-12 English learners, English proficiency placement tests may also include global speaking and listening rubrics (see http://www.doe.in.gov/sites/default/files/elme/el-guidebook-10-29-13.pdf for an example).

REFERENCES

Archibald, J. 1998. *Second Language Phonology*, Philadelphia: John Benjamins.

Benrabah, M. 1997. Word stress – a source of unintelligibility in English. *IRAL* 35(3): 157–165.

Bowen, J.D. 1972. Contextualizing pronunciation practice in the ESOL classroom. *TESOL Quarterly* 6: 83–94.

Bradlow, A., Akahane-Yamada, R., Pisoni, D., and Tohkura, Y. 1999. Training Japanese listeners to identify English /r/and /l/: long-term retention of learning in perception and production. *Perception and Psychophysics* 61: 977–985.

Canale, M. and Swain, M. 1980. Theoretical bases of communicative approaches to

second language teaching and testing. *Applied Linguistics* 1: 1–47.

Celce-Murcia, M., Brinton, D., Goodwin, J., and Griner, B. 2010. *Teaching Pronunciation*, New York: Cambridge University Press.

Counselman, D. 2010. *Improving Pronunciation Instruction in the Second Language Classroom.* Retrieved from: http://search.proquest.com/docview/898324856?accountid=14707. (898324856; ED522787).

Darcy, I., Ewert, D., and Lidster, R. 2012. Bringing pronuciation instruction back into the classroom: an ESL Teachers' Pronunciation "ToolBox". In: *Proceedings of the 3rd Pronunciation in Second Language Learning and Teaching Conference*, J. Levis and K. LeVelle (eds.), 93–108, Ames: Iowa State University.

Derwing, T.M., Munro, M.M., and Wiebe, G. 1998. Evidence in favor of a broad framework for pronunciation instruction. *Language Learning* 48: 393–410.

Derwing, T.M. and Rossiter, M.J. 2002. ESL learners' perceptions of their pronunciation needs and strategies. *System* 30(2): 155–166.

Doughty, C. and Williams, J. 1998. Pedagogical choices in focus on form. In: *Focus on Form in Classroom Second Language Acquisition*, C. Doughty and J. Williams (eds.), Cambridge: Cambridge University Press.

Ellis, R., Loewen, S., and Basturkmen, H. 2006. Disentangling focus on form. A response to Sheen and O'Neill (2005). *Applied Linguistics* 27: 135–141.

Escudero, P., Hayes-Harb, R., and Mitterer, H. 2008. Novel second-language words and asymmetric lexical access. *Journal of Phonetics* 36: 345–360.

Field, J. 2005. Intelligibility and the listener: the role of lexical stress. *TESOL Quarterly* 39: 399–423.

Field, J. 2014. Pronunciation teaching has to establish in the minds of language learners a set of distinct consonant and vowel sounds. In: *Pronunciation Myths: Applying Second Language Research to Classroom Teaching*, L. Grant (ed.), 80–106, Ann Arbor, MI: University of Michigan Press.

Foote, J.A., Holtby, A.K., and Derwing, T.M. 2011. Survey of the teaching of pronunciation in adult ESL Programs in Canada, 2010. *TESOL Journal Canada* 20: 1–22.

Gatbonton, E. and Segalowitz, N. 1988. Creative automatization: principles for promoting fluency within a communicative framework. *TESOL Quarterly* 22: 473–492.

Gilbert, J. 1995. Pronunciation practice as an aid to listening comprehension. In: *A Guide for the Teaching of Second Language Learning*, D.J. Mendelsohn and J. Rubin (eds.), 97–111, San Diego, CA: Dominie Press.

Gilbert, J. 2001 *Clear Speech from the Start: Basic Pronunciation and Listening Comprehension in North American English*, New York: Cambridge University Press.

Jiang, N. 2007. Selective integration of linguistic knowledge in adult second language learning. *Language Learning* 57: 1–33.

Kachru, B. 2005. Teaching World Englishes. *Indian Journal of Applied Linguistics* 15: 85–95.

Lardiere, D. 2003. Revisiting the comparative fallacy: a reply to Lakshmanan and Selinker, 2001. *Second Language Research* 19: 129–143.

Levis, J.M. and Barriuso, T.A. 2012. Nonnative speakers' pronunciation errors in spoken and read English. In: *Proceedings of the 3rd Pronunciation in Second Language Learning and Teaching Conference*, J. Levis and K. LeVelle (eds.), 187–194, Ames: Iowa State University.

Levis, J.M. and Grant, L. 2003. Integrating pronunciation into ESL/EFL classrooms. *TESOL Journal* 12: 13–19.

Loschky, L. and Bley-Vroman, R. 1993. Grammar and task-based methodology. In: *Tasks and Language Learning: Integrating Theory and Practice*, G. Crookes and S. Gass (eds.), 123–167, Clevedon, UK: Multilingual Matters.

Major, R.C. 1987. A model for interlanguage. In: *Interlanguage Phonology: The Acquisition of a Second Language Sound System*, G. Ioup and S.H. Weinberger (eds.), Cambridge: Newbury House Publishers.

Markham, D. 1997. *Phonetic Imitation, Accent and the Learner*, Lund: Lund University Press.

McCrocklin, S. 2012. Effect of audio vs. video listening exercises on aural discrimination of vowels. In: *Proceedings of the 3rd Pronunciation in Second Language Learning and Teaching Conference*, J. Levis and K. LeVelle (eds.), 178–186, Ames: Iowa State University.

Murphy, J.M. 1991. Oral communication in TESOL: integrating speaking, listening, and pronunciation. *TESOL Quarterly* 25: 51–75.

Murphy, J.M. and Kandil, M. 2004. Word-level stress patterns in the academic word list. *System* 32: 61–74.

Pennington, M.C. and Ellis, N.C. 2000. Cantonese speakers' memory for English sentences with prosodic cues. *The Modern Language Journal* 84: 372–389.

Pica, T., Kanagy, R., and Falodun, J. 1993. Choosing and using communication tasks for second language instruction. In: *Tasks and Language Learning: Integrating Theory and Practice*, G. Crookes and S. Gass (eds.), 9–34, Clevedon, UK: Multilingual Matters.

Prator, C. 1971. Phonetics vs. phonemics in the ESL classroom: When is allophonic accuracy important? *TESOL Quarterly* 5: 61–72.

Reed, M. 2012. The effect of metacognitive feedback on second language morphophonoly. In: *Proceedings of the 3rd Pronunciation in Second Language Learning and Teaching Conference,* J. Levis and K. LeVelle (eds.), September 2011, 168–177, Ames: Iowa State University.

Reed, M. and Michaud, C. 2005. *Sound Concepts: An Integrated Pronunciation Course*, New York: McGraw-Hill.

Reed, M. and Michaud, C. 2011. An integrated approach to pronunciation: listening comprehension and intelligibility in theory and practice. In: *Proceedings of the 2nd Pronunciation in Second Language Learning and Teaching Conference*, J. Levis and K. LeVelle (eds.), 95–104, Ames: Iowa State University.

Robinson, P. 2001. Task complexity, task difficulty, and task production: exploring interactions in a componential framework. *Applied Psycholinguistics* 21: 27–57.

Rvachew, S., Nowak, M., and Cloutier, G. 2004. Effect of phonemic perception training on the speech production and phonological awareness skills of children with expressive phonological delay. *American Journal of Speech–Language Pathology* 13: 250–263.

Saito, K. and Lyster, R. 2012. Effects of form-focused instruction and corrective feedback on L2 pronunciation development of /ɹ/ by Japanese learners of English. *Language Learning* 62: 595–633.

Segalowitz, N. and Hulstijn, J. 2005. Automaticity in bilingualism and second language learning. In: *Handbook of Bilingualism. Psycholinguistic Approaches*, J.F. Kroll and A. De Groot (eds.), 371–388, New York: Oxford University Press.

Sicola, L. 2009. *"No, They Won't 'Just Sound Like Each Other'": NNS–NNS Negotiated Interaction and Attention to Phonological Form on Targeted L2 Pronunciation Tasks*, Frankfurt: Peter Lang.

Trofimovich, P. and Gatbonton, E. 2006. Repetition and focus on form in processing L2 Spanish words: implications for pronunciation instruction. *The Modern Language Journal* 90: 519–535.

van Loon, J. 2002. Improving pronunciation of adult ESL students. *TESOL Canada Journal* 20: 83–88.

Wang, W., Jongman, A., and Sereno, J.A. 2003. Acoustic and perceptual evaluations of Mandarin tone productions before and after perceptual training. *Journal of the Acoustical Society of America* 113: 1033–1043.

Williams, J. 2003. *Preparing to Teach Writing: Research, Theory and Practice*. 3rd edition, Mahwah, NJ: Lawrence Erlbaum Associates.

Willis, D. and Willis, J. 2007. *Doing Task-Based Teaching*, Oxford: Oxford University Press.

Yang, Y. and Lyster, R. 2010. Effects of form-focused practice and feedback on Chinese EFL learners' acquisition of regular and irregular past tense forms. *Studies in Second Language Acquisition* 32: 235–263.

Zielinski, B.W. 2008. The listener: no longer the silent partner in reduced intelligibility. *System* 36: 69–84.

27 Using Orthography to Teach Pronunciation

WAYNE B. DICKERSON

Introduction

It is hard to find a more stinging indictment of English spelling than that which opens Professor Mont Follick's *Case for Spelling Reform* (1965: 1), "Our present spelling system is just a chaotic concoction of oddities without order and cohesion." This is not a reformer's hyperbole; the sentiment is as widespread, deeply rooted, and profoundly felt across the English speaking community today as it was 50 years ago. Word pairs like *to* and *go*, *few* and *sew*, *gauze* and *gauge*, illustrating problematic symbol–sound associations, only reinforce the skepticism among laymen and teachers alike that anything good can come of our present spelling system.

In the face of such withering criticism and public disdain for how we spell words, it might seem like a fool's errand to suggest that English spelling can actually be a useful ally in our job of helping ESL/EFL learners improve their oral English skills. Even so, that is the intent of this chapter. It asks those who would discredit our whole writing system because of apparently anomalous spellings to suspend judgment and take a dispassionate look at how our orthography actually goes about representing English sounds. They would likely be amazed at how much valuable information can be gleaned about spoken English from its "defective" spellings.

While Morley (1994), citing Dickerson (1989, now 2004), called attending to sound–spelling relationships a "major shift in the instructional focus of ESL programs" (1994: 65), this shift has been very slow to materialize. Hahn and Dickerson (1999) show advanced learners how to use spelling to predict the major stress of words of any length. Gilbert (2001), drawing on patterns from vowel phonics that apply to one-syllable words, offers low-level learners some "… simple and efficient spelling rules to guess how a word is pronounced" (2001: x).

Much more than this can be done with spelling, as this chapter will illustrate. We begin by attempting to dispel a myth about English spelling, namely, that it

The Handbook of English Pronunciation, First Edition. Edited by Marnie Reed and John M. Levis.

represents how we pronounce words. Then we explore the areas in which ESL/EFL learners and their teachers can benefit from spelling guidance.

Representing English sounds

Conventional wisdom says that an ideal orthography should match each vowel and consonant sound (each phoneme) in the language uniquely to one letter. Russian, Spanish, Finnish, Serbian, and other languages approach this ideal. The belief that English should do the same underlies the endless roasting of English spelling for its seemingly "chaotic concoction of oddities without order and cohesion".

If English spelling were built on the one-sound-to-one-symbol principle, our way of spelling words would deserve every criticism leveled at it. However, connecting sounds and letters one-to-one is *not* its dominant principle. Instead of representing sound directly, English spelling attempts to represent meaning directly. Without pronouncing these words, how do readers recognize the past-tense ending in *it appealed, she recited, he politicked?* The uniform *-ed* spelling tells us immediately. It makes no difference to readers that the *-ed* is pronounced /d/, /əd/, and /t/ respectively. By putting meaning first and by spelling related words and word parts similarly, the spelling system helps readers grasp by sight the semantic connection between, for example, *appeal* and *appellation*, *recite* and *recitation*, *politick* and *politician*. The fact that the *ap-* prefix of *appeal* and *appellation* is pronounced /ə/ and /æ/ respectively does not bother readers. We are not disturbed that the *-cit-* root is /ay/ in *recite* but /ə/ in *recitation*. Nor do readers hesitate over the /k/ and /ʃ/ pronunciations of the *c* of *politick* and *politician* respectively. In their attention to meaning, native readers are largely oblivious to the fact that the same letters are used for different sounds.

While English represents meaning directly, by-passing sound, it represents sound indirectly. The principle in play is this: whatever is predictable by rule is unwritten. To extract sound from spelling, one has to know the rule linking letters to sounds. For example, the variant sounds of *-ed* are unwritten because readers know that after a *t* or *d* the *-ed* will be pronounced as /əd/; after a vowel or voiced consonant other than *d*, *-ed* will be pronounced as /d/; and after a voiceless consonant other than *t*, *-ed* will be pronounced as /t/. The regular alternation of full vowels and the reduced vowel /ə/ in the first two syllables of these "flawed" words *appeal – appellation, recite – recitation,* and *politick – politician* also goes unwritten because the vowel alternation exactly matches the predictable alternation of stress: ˘ ´ versus ` ˘ in these words. The divergent pronunciations of *c* are not coded in spelling because readers know that *c* before *i* and another vowel letter (e.g., *-ian*) is predictably pronounced as /ʃ/ but as /k/ when the *c* comes before another consonant letter and before the letters *a*, *o*, *u* (e.g., *crack*, *cat*, *cot*, *cut*) (Dickerson 1994; Kreidler 1972; O'Neil 1980).

By representing meaning directly and sound indirectly, our spellings make it inherently easier to extract meaning from written words than to extract sound.

To discover sounds in spelled words, learners *must use rules* like those above. There is no alternative to rules; they are part of the essential nature of our spelling system. Native decoders know some of these rules by virtue of being native speakers. They have learned other rules just like non-native decoders must learn them. Unaware of the rule-mediated connection between spelling and sound and the absolute necessity of rules, textbook writers and teachers often hesitate to offer learners the rules needed to derive pronunciations from spelling. Wishing for a direct spelling–sound link, textbook writers and teachers tend to reject rules, even in a learnable form, as too complex for students. As a result, denied access to rules, their students are effectively denied access to sound via spelling, which could be a life-long resource for them.

While challenging, rule learning is not beyond the ESL/EFL student's capability when they use *learner rules*, specially designed formulae (illustrated below) that take into account the learner's limitations (Dickerson 2012). Fixated on anomalies, detractors of spelling often fail to appreciate that English spelling does an amazingly good job preserving all the structural information necessary for learner rules to work well – prefix, stem, and suffix identities, syllable count and syllable structure, and even cues to guide the selection of symbols for consonant and vowel segments. Even "sight words", which defy decoding strategies in some respects (Otto and Chester 1972) and are parceled out to elementary school children to memorize, are not chaotic. Except for words like *one*, *once*, those with silent letters like *talk*, *could* and those with anomalous *gh* spellings, even sight words have consonant letters where consonant sounds are and vowel letters where vowel sounds are.

An assessment of how well-formed English spelling is depends on the yardstick used. Clearly direct symbol-sound connections will not do. Instead, the assessment tool must gauge English spelling according to principles at work in the system: (a) How well does spelling preserve visual evidence of the semantic connections among related words (meaning-first principle)? (b) Is the information in spelling rich enough that learner rules can apply to generate conventional pronunciations (sound-second principle)? This is precisely the metric that Chomsky and Halle used when concluding that standard orthography is remarkably close to an ideal representation of English words (1968: 48–49, 96, 184n).

With that endorsement of our spelling system, we turn now to the role that English spelling can play in the learner's developing sound system and to examples of learner rules that can make available valuable pronunciation clues.

Orthography for prediction

Attention to spelling can be of use in pronunciation teaching and learning to the extent that it helps teachers and learners realize goals they value. We understand of course that good **production** skills underlie good communication. Equally important is an ability to hear what is said to us so that we can interpret the messages sent. Good **perception** skills are likewise essential. Less widely appreciated, but no less fundamental to communication, is the ability to make good judgments *before*

speaking about what to say in each area of production. Good **prediction** skills make possible good production and perception. These three skills – prediction, production, perception – are what we call the **3Ps** (Dickerson 2004: Unit 1, 8). Since we value these skills as fundamental to communication, and since our basic objective in teaching pronunciation is to help learners develop intelligible oral communication, we take the 3Ps as our pronunciation-teaching goals.

Of these three goals, the use of orthography serves prediction most directly and, through prediction, it serves production and perception. The use of orthography for prediction has its place principally in a strategy chain we call *covert rehearsal*, in which learners privately inspect their oral utterances, evaluate them against rules they know and models they have learned, correct them, and then practise their corrections until they can say them fluently and accurately (Dickerson 2000). To use orthography in this way, learners need to know useful patterns they can apply to spelled words. They also need to learn how to use this strategy chain effectively. The effort to equip learners with internal resources – providing rules, models, and practice with the strategy chain – is important because long-term pronunciation improvements can result from using these resources (Sardegna 2009).

Predicting consonant choice

If there is one area of phonology that is iconically identified with pronunciation teaching it is the area of segmentals – the vowels and consonants that make up each word. Their claim to fame is that they do the work of distinguishing one word from another. They keep *useful* from sounding like *youthful* and *mislaid* from sounding like *misled*. Segmentals that do this kind of work are called **phonemes**. That is why phonemes are so central to pronunciation teaching.

Predicting consonant phonemes via orthography is different from predicting vowel phonemes. That is because consonant choice is not so tightly bound to the stress of a word as vowel choice is. Even so, decoding consonant letters and letter combinations is not straightforward.

Learners who anticipate being able to judge consonant sounds directly from consonant spellings will be disappointed. Only half of the consonant letters in the alphabet point unambiguously to a single consonant phoneme: *b*, *f*, *j*, *k*, *m*, *p*, *q*, *r*, *v*, and *z*. The other half have no such immediate connection to a consonant phoneme. To these we can add letter combinations such as *ch*, *sch*, *gh*, *sh*, *th*, *ng*, *ps*, and *pt*. Only *ph*, *wr*, *mb*, *mn*, *pn*, and *kn* reliably point to only one phoneme each.

While there are some direct symbol-to-sound connections among consonants, the great majority of letter-to-sound connections are indirect, requiring the use of rules to determine the phonemic value of a letter or letter combination. That is, most letters are busy implementing the meaning-first principle, identifying for the eye of the reader the semantic relatedness of words such as *political* and *politician*. For readers to extract a pronunciation from such spellings, the visual shape of the words must preserve enough information that the rules of the sound-second principle can generate a pronunciation successfully.

Clues to the sound value of a graphic unit (letter and letter combinations) are to be found in the environment surrounding the letter – its neighboring letters, nearby endings, its position in a word, degrees of stress on adjacent vowels, or a combination of these clues. An analysis of the *c* in words like *politician*, *political*; *electrician*, *electricity*, *electric* reveals these regularities, with the most specific rule given first and the most general given last. The rules form an ordered set:

Sound	*Environment*	*Examples*
/ʃ/	c+iV-ending	iV-endings are strings like *-ia*, *-ion*, *-ial*, *-ious*, *-ient.* e.g., *aca**c**ia*, *suspi**c**ion*, *offi**c**ial*, *gra**c**ious*, *effi**c**ient*
/s/	ce/i/y	The / means "replace the letter on the left with the letter on the right", which gives us *ce*, *ci*, *cy*, e.g., ***c**eiling*, *pea**c**e*, ***c**ity*, *defi**c**it*, ***c**ypress*, *mer**c**y*
/k/	$c_{elsewhere}$	"Elsewhere" means "not in the above environments", e.g., ***c**all*, *sta**c**k*, *a**c**tive*, *traffi**c***

The sound-second principle says that a graphic unit in a word context predicts its sound value. Since this is the way our orthography works, we have designed prediction rules for learners around this principle. Based on the analysis above, learner rules for the consonant letter *c* are the following, to be used in this order:

A graphic unit in context	*predicts*	*its sound value*
c+iV	=	/ʃ/
ce/i/y	=	/s/
c_{ew}	=	/k/

Consonant prediction patterns such as these are also written to conform to the characteristics of a good learner rule (Dickerson 2012). One feature of a good learner rule is that it is stated succinctly enough that it can be practised easily in written exercises, ideally in coordination with articulatory work on one of the key segmental targets. For example, the above patterns can be presented when working on palatal consonants, thereby joining prediction to production and perception work.

To illustrate learner rules for a consonant letter combination, we can look at the interpretation of the *th* spelling, troublesome for learners of English trying to articulate the difference, and even for native speakers of English trying to tell the difference between /θ/ and /ð/:

A graphic unit in context	*predicts*	*its sound value*
thV^{f}	=	/ð/
thern/·	=	/ð/
V/rth+E	=	/ð/
th_{ew}	=	/θ/

The consonant eth (/ð/) occurs almost exclusively among native Anglo-Saxon words; words borrowed from Greek and elsewhere entered the language with /θ/. Despite the borrowings, the environments are sufficiently distinct that only a dozen out of about 800 *th* words cannot be predicted by the rules above. The first rule says that when *th* is followed by a vowel letter in a function word ([f]), the phoneme value of *th* is /ð/, as in *the*, *this*, *them*, *although.* The second rule says that when we see a *thern* string or a *ther·* string, the *th* should be pronounced as /ð/, as in *northern*, *farther*, *bothered*. (The · symbol stands for end of word or before an ending like *-e*, *-ed*, *-ing*.) The third rule applies to V*th* or *rth* followed by an ending (E) such as *-e*, *-ed*, *-ing*. Again, the predicted phoneme value of *th* is /ð/, as in *farthing*, *bathe*, *seethed.* The last rule tells us that every other instance of *th* should be pronounced as /θ/. These patterns can be practised in written exercises when teaching /θ/ and /ð/ (Dickerson 2006).

Predicting major word stress

Without accent marks in standard written English, or a uniform stress-placement rule, an English text tells us nothing directly about where the stressed and unstressed vowels are. For example, nothing in the words *colony* and *colonial* indicates that the first two *o* letters in *colony* are stressed and unstressed and that the first two *o* letters of *colonial* are unstressed and stressed. Consistent with the nature of English orthography, the stress of a word can be ascertained only indirectly by rule.

An indirect approach to word stress should not deter us; word stress is too important to be ignored. It is a subsystem that supports the entire structure of phonology. Fortunately, it is also a part of phonology that can be predicted using learner rules that apply to standard orthography.

Wherever the major stress of a word falls on a polysyllabic word, it creates one of three possible word-rhythm patterns in English relative to the peak: peak-valley (e.g., *cómplicated*), valley-peak-valley (e.g., *persecútion*), or valley-peak (e.g., *represént*). Major stress on the right syllable, creating the right rhythm, makes a spoken word intelligible. On the wrong syllable, creating an unexpected rhythm, it may obscure its meaning entirely (Field 2005).

Most importantly for interpersonal communication, major word stress has the potential to contribute a meaningful peak to the discourse, thereby adding substantially to the listener's understanding of the message. A peak has the power to signal, however, largely because it contrasts with surrounding valley syllables. That is, for maximal effect, it is not enough for teachers and learners to focus on peaks and ignore nearby valleys. Both are equally important, which is why contrast is such a fundamental feature of oral communication.

Understanding the importance of word stress, generations of ESL/EFL teachers and textbook writers have tried to help. With no simple way to determine the location of a word's major stress, they have offered a variety of partial solutions: citing statistical guidance (Prator and Robinett 1972), suggesting that the practice

of words with particular patterns will help the pattern "rub off" (Trager and Henderson 1957), offering endings where stress can be predicted reliably, e.g., *-ion*, etc. (Woods 1979). While helpful, none of these approaches adds up to a systematic or comprehensive way to stress all, or even a useful majority of, the words in English.

The only fully developed word-stress prediction system yet available to ESL/EFL learners is that found in Dickerson (2004), which was strongly influenced by the research of Chomsky and Halle (1968). Others have also worked profitably in this arena (Guierre 1984; Teschner and Whitley 2004). The broad outlines of the prediction system in Dickerson (2004) are presented to illustrate what can be done.

To understand the rule system, we need to recognize that, regardless of the length of a polysyllabic word, it will carry its major stress on only one of two syllables, either on the Key Syllable or on the Left Syllable. These two syllables can be identified unambiguously in spelling terms for every polysyllabic word. For example, the Key Syllable (as underlined) is immediately to the left of particular endings, e.g., *punit(ive*, *regul(atory*, *compass(ionately*. Sometimes, if there is no ending, the Key Syllable is the last syllable, e.g., *underdevelop*, *disregard*, or the next-to-the-last syllable, e.g., *maverick*, *astronaut*, depending on the part-of-speech. The Left Syllable is always immediately to the left of the Key Syllable. The fact that there are only two candidates for major stress – the Key and Left Syllables – and that both are easy to locate in any word hugely reduces the chances of putting the stress on the wrong syllable. The role of the four word-stress rules in Dickerson (2004) is to reduce those chances to almost zero.

The four word-stress rules start at the Key Syllable; two rules focus on the Left Syllable and two focus on the Key Syllable. Of the two rules focusing on the Left Syllable, one places stress directly on that syllable (Left Stress Rule). The other examines the composition of the Left Syllable (whether or not any part of a prefix is present) to determine whether the stress should go on the Key Syllable or on the Left Syllable (Prefix Stress Rule). Of the two rules focusing on the Key Syllable, one is designed to place stress directly on that syllable (Key Stress Rule). The other examines the composition of the Key Syllable (its syllable structure) to determine whether the stress should go on the Key Syllable or on the Left Syllable (V/VC Stress Rule). The focus of each rule and its way of assigning stress are depicted in the following summary where SR stands for "stress rule". Among the four stress rules, the major stress of every English word is accounted for with few exceptions.

		Focus: Left	Focus: Key
Method	Direct	Left SR	Key SR
	Evaluate	Prefix SR	V/VC SR

Within this structure, the learner's prediction task involves answering three questions:

Q1 Which rule applies to a word?
Q2 Where is the Key Syllable?
Q3 Where does the rule place major stress, on the Key or Left Syllable?

Years of empirical research have answered the first question. The Key Syllable is defined by each rule. As indicated, the position of the Left Syllable is derived from the position of the Key Syllable. The rule then applies to place the major stress on the Key or Left Syllable.

To provide a sense of how the rules actually work, and to show that the rules are not difficult to use, we will illustrate the Left Stress Rule and the V/VC Stress Rule as each applies to a narrowly defined word group. One rule focuses on the Left Syllable and the other on the Key Syllable. One rule places stress directly and the other places stress by evaluating the composition of a syllable.

The Left Stress Rule applies to words that end in *-ate* and derivatives (*-ates*, *-ated*, *-ating*, *-ator*). These words have two or more syllables left of the ending (Q1). The Key Syllable (underlined in the examples) is immediately to the left of the ending (Q2). The Left Syllable is left of the Key Syllable. The Left Stress Rule places stress on the Left Syllable (Q3):

Examples of Stress Left: *cónfisc*(*ated*, *indiscrímin*(*ate*, *démonstr*(*ator*, *commúnic*(*ating*

The V/VC Stress Rule applies to words that end in *-ous* (Q1). The Key Syllable (underlined in the examples) is immediately to the left of the ending (Q2). The Left Syllable is left of the Key Syllable. The V/VC Stress Rule places stress by evaluating the Key Syllable: Is it spelled with a single vowel letter (V) or a single vowel letter followed by a single consonant letter (VC)? If so, stress the Left Syllable. If not, stress the Key Syllable (Q3):

Examples of Stress Left: *impétu*(*ous*, *ambígu*(*ous*, *húmor*(*ous*, *anónym*(*ous*

Examples of Stress Key: *treménd*(*ous*, *momént*(*ous*, *polymórph*(*ous*, *disástr*(*ous*

This word-stress system empowers learners with the internal resources to stress tens of thousands of words with great accuracy. Stress exceptions in each word group are usually under 1%. Even so, the rules must be used selectively because time allotted for attention to pronunciation is always limited. For all learners, including those who cannot take full advantage of this resource, what are the most important take-aways about English word stress? These stand out.

1. The major stress of a word is predictable.
2. The major stress will fall on the Key or Left Syllable.
3. The major stress creates one of three rhythms in every polysyllabic word. (Practice identifying the rhythm pattern of polysyllabic words is worthwhile.)

4. The location of the Key Syllable is predictable, usually just to the left of an ending. (Time spent finding the Key Syllable in different word groups is time well spent.)
5. The Left Syllable is always immediately to the left of the Key Syllable.
6. Finding the Key and Left Syllables can limit stress guessing. (Hearing stressed words and saying words with the stressed vowel marked can improve guessing.)
7. Stress rules are so straightforward that post-puberty learners can learn to use them even on their own if they wish. (Well-structured materials can help, e.g., Hahn and Dickerson 1999.)

Predicting major-stressed vowels

Efforts to predict vowel sounds from spelling have been part of pronunciation instruction for many decades (Prator 1951; Vernick and Nesgoda, 1980; Guierre 1984; Gilbert 2001). That in itself is a testament to the fact that there are useful regularities in how spellings point to sounds.

Vowel prediction, like consonant and word-stress prediction, follows the sound-second principle: *A graphic unit in context predicts its sound value*. A graphic unit is a single vowel letter or a letter combination. Its context in a word must include those factors that are relevant to the language. To do a good job, a vowel prediction pattern must take into account word stress, neighboring letters, and position in a word.

A learner rule incorporates all three conditions. On the left of the pattern, left of the = mark, is a vowel letter or a general stand-in for a vowel letter, V, in its relevant context, and on the right is the predicted vowel phoneme or a vowel quality. Two vowel patterns illustrate the presence of the three essential ingredients of context (Dickerson 1980):

A graphic unit in contex	*predicts*	*its sound value*
úC←	=	tense
V́C←	=	lax

For both, the syllable in question carries major stress, as determined beforehand by a stress rule. This syllable consists of a single vowel letter followed by a single consonant letter. The first pattern applies exclusively to the letter *u*; the second case is not specific to particular vowel letters. The left-pointing arrow designates the syllable in each case as the Left Syllable. On the right, the first pattern predicts a tense vowel, namely, /uw/. The second pattern reliably predicts a lax vowel. These are ordered rules. That is, the first and most specific rule filters out *u* cases; the second, more general, rule applies to all other vowel letters. The first rule tells us that the uC Left Syllables in *púnitive*, *communicating*, and *humorous* (all mentioned above) should be pronounced as /uw/. The second rule tells us why the vowel letter in the stressed VC Left Syllables of *colony*, *demonstrator*, and *ambiguous*

(all mentioned above) have lax vowels. A full presentation of vowel rules such as these, designed for ESL/EFL learners, is given in Dickerson (2004), including the tool needed to translate "tense" or "lax" into a specific vowel prediction (see also Dickerson 2012).

After assigning the major stress to one syllable of a word, predicting vowels left of the major stress is a much easier task. Left of the major stress, we can predict the stress and vowel quality simultaneously. The complete system is presented in Dickerson (2004).

Predicting compression

Good rhythm when speaking English promotes intelligibility. Rhythm, however, is not a single phenomenon but a collection of phenomena that can be grouped into two meaning-based categories – contrast and compression. Contrast is the difference between peaks – longer, louder, and higher-pitched vowels – and valleys – shorter, quieter, and lower-pitched vowels – in a phrase. Compression refers to the many ways we abbreviate valley syllables across a phrase – those in function words and content words alike. In focus here is the feature of compression.

We use peaks to highlight words that carry more significant meaning (typically content words and certain function words) and valleys for words of less significance. In a typical phrase of one or two peaks, the majority of syllables are in valleys (Bolinger 1986: 47–48). English speakers not only highlight the peaks by contrasting them with surrounding quieter, briefer, and lower-pitched valleys but they also hurry the less significant valley words along. They use a variety of devices to minimize the vowels and consonants of these syllables. The effect is to draw the peaks closer together, as listeners expect. By meeting this expectation, speakers enhance their intelligibility. We can capture the compression devices we use, in the order we use them in speech, in this convenient acronym: NATRL – native assimilation, trimming, reduction, and linking.

NATRL devices are accessible to learners in part because English orthography faithfully represents the structure of spoken syllables, using consonant letters for consonant sounds and vowel letters where vowel sounds belong. Compression devices are also accessible because learners can easily understand the rules that apply to these spellings. The combination of a rich orthography, learner-oriented rules, and the high-value feature of good compression make NATRL devices an ideal starting point for using ordinary spelling to improve the clarity of learners' spoken language. The most important native-English compression devices are the following (Hahn and Dickerson 1999):

A: In American English, palatal assimilation compresses two segments, an alveolar nonsonorant (/t/, /d/, /s/, /z/) and a palatal glide /y/ (at the start of *you*, *your*, *yourself*) into a single palatal segment /ʧ/, /ʤ/, /ʃ/, and /ʒ/ respectively.

/ʧ/ Do it_yourself! He guessed_your secret.
/ʤ/ You included_yourself. Would_you help me?
/ʃ/ Can you dress_yourself? Try to trace_your roots.
/ʒ/ It taxes_your brain. Whatever pleases_you!

T: Trimming is the complete loss of a vowel or consonant segment. Five types of trimming save valley time. In describing trimming, we use an apostrophe to mark the position of the loss.

Loss of /t/ and /d/ from Ct and Cd clusters. This loss happens when a consonant (but not *w*, *h*, *y,* or *r*) follows the cluster. All Ct and Cd clusters are affected by trimming except *lt*, *nt*, *rt, rd*, and *r-ed*. Most examples occur at word boundaries, e.g., *mos' people*, *kep' singing*, *mov'(ed) quickly*. However, /t/ and /d/ will also be lost between word parts, e.g., *han'some*, *cos'ly*, *enac'ment*.

Loss of consonants and vowels from contractions. Contractions commonly trim some portion of eleven function words (*am*, *is*, *has*, *are*, *did*, *had*, *would*, *have*, *will*, *us*, *not)*: *I'm*, *she's here*, *she's gone*, *you're*, *he'd go*, *he'd gone*, *where'd he go? we've*, *they'll*, *let's*, *can't*. The apostrophe indicates that a loss has occurred. It does not identify what has been lost – a vowel sound, a consonant sound, or both. Nor does it suggest how to pronounce the remainder. Depending on the word it is attached to, a contraction may have two or three different pronunciations.

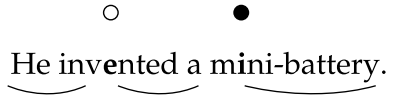
H-loss from he, him, his, her, have, has, had. When not at the start of a phrase nor under primary stress, the /h/ of these seven function words will drop away. An eighth instance of h-loss happened to the ancient form of *them*, *hem*, and continues to the present as *'em* (Pyles 1964: 334), e.g., *Tell 'im about it. I should 'ave warned you. Go get 'em!*

Vowel loss with a syllabic consonant. In a string of two syllables where the first is stressed and the second is unstressed, the vowel of the second syllable will drop away most commonly when the first syllable ends with /t/ or /d/ and the first consonant after the next vowel is /l/ or /n/. The /l/ or /n/ that remains carries the beat; it becomes the center of the syllable ("syllabic") like the vowel that has been lost, e.g., *met'l*, *id'l*, *sent'nce*, *gard'n*.

Vowel loss without a syllabic consonant. In a string of three syllables where the first is stressed and the next two are unstressed, the middle vowel will drop away most often before a single /n/, /l/, or /r/ consonant, e.g., *comp'ny*, *fam'ly*, *ev'ry*.

R: Reduction preserves the segment but shrinks its size. The most important reduction is vowel reduction. While all valley vowels are reduced in size, the most common reduced vowel is the schwa [ə]. Reduced valley vowels are so important that they are required; all the other NATRL devices are optional. They are important because they alone serve two functions. They contrast with peaks to make peaks stand out. For example, in the following sentence, *-vent-* and *min-* stand out in part because they are surrounded by valley vowels. Reduced vowels also do more to speed up valley syllables than any of the

other devices because there is a reduced vowel in almost every valley syllable. In this example, there are eight valley vowels to two peak vowels.



In its work on behalf of compression, vowel reduction gets help from consonant reduction in American English. Oral and nasal flaps reduce the size of /t/, /d/, /n/, and /nt/ to a fraction of their nonflapped duration. This happens mostly when these segments come between stressed and unstressed vowels or between two unstressed vowels, as we hear in these parts of the sentence above: *-ve**nt**ed*, *-e**d** a*, *-i**n**i*, *-a**tt**e-*.

L: While all adjacent segments in a phrase are close together, only a few interact with each other to shorten their overall articulation time. We refer to four cases that interact this way as linking.

$C\smile C_{same}$	Linking between identical continuants simply continues the first consonant a little longer, e.g., *yes‿sir!*, *a rough‿few days*, *the same‿moment.* Linking identical stops involves holding the stoppage a little longer, e.g., *rob‿banks*, *not‿talking*, *a big‿group.*
$C_{st}\smile C_{st/af/nas}$	When a stop is adjacent to a different stop, an affricate, or a nasal, the air of the first stop is not released until the tongue shifts to the new position. This happens between words, as in *back‿pain*, *good‿morning*. It also happens within words, as in *elec‿tive*, *ab‿normal*, *mag‿nificent*.
C‿V/w y r/	The consonant at the end of a word seems to attach itself to the vowel, /w/, /y/, or /r/ at the start of the next word, e.g., *ask‿about*, *closed‿it*, *the best‿way*, *some‿years ago*, *a pack‿rat*.
$V\overset{w}{\underset{y}{\smile}}V$	Except for word-final schwa, all other word-final vowels use their off-glides (/-w/, /-y/) as a bridge to a vowel-initial word, e.g., *go‿on*, *fly‿over*. The glide moves to the next syllable. This process is also seen inside words where two vowel sounds juxtapose, e.g., *the‿ology*, *mightier*, *co-author*, *tuition*.

A concentration on linking is particularly important for students who insert a glottal stop (a "throat stop") [ʔ] before every vowel-initial word and for those who insert a schwa [ə] between the consonant end of one word and the consonant start of another. Without the help of linking, their speech stream sounds choppy and is distracting to the listener. Fortunately for such students,

ordinary spelling represents the relevant segments accurately enough that they are able to identify types of linking with a high level of accuracy.

Learners have the advantage that most of these devices are presented in modern pronunciation textbooks (Weinstein 2001). Good discussions are also available in teachers' guides (e.g., Celce-Murcia, Brinton, and Goodwin 2010: 163–184). While their use is encouraged in order to sound more natural and friendly, they are worth teaching, especially because they meet listeners' expectations.

Predicting suffix forms

In our prefix-stem-suffix language, endings abound. How we say them makes a difference, particularly if they carry significant grammatical information. As we have come to appreciate, our spelling system preserves each ending in a uniform graphic shape regardless of how it is pronounced. Fortunately, the rules needed to give them a pronunciation are not complex.

The most important endings for learners are *-ed* and *-s* because of the information load they carry. The *-ed* marks the past tense and past participle verb, and derivative nouns and adjectives: *They dedicated the park*, *She has dedicated her life to service*, *These are the dedicated*, *They are so dedicated*.

The *-s* endings, *-'s* genitive, and *-'s* contractions are used to make nouns plural and possessive, verbs third-person present tense, and to shorten *is*, *has*, and *us*, e.g., *my sisters*, *my brother's wife*, *she works*, *she's working*, *she's been living*, *let's go*.

Most of the *-ed*, *-s*, and *-'s* meaning units have potentially three forms each, /t/, /d/, /əd/ and /s/, /z/, /əz/. We know, however, that the exact voicing of the single-sound variant is not crucial for intelligibility as long as the voiced or voiceless variant is present. By removing the voicing decision (/t/ versus /d/ and /s/ versus /z/), we can simplify the three-way decision to a two-way decision and improve the learner's accuracy (Dickerson 1990).

The decision procedure uses orthography in place of sound as the basis for prediction. The rules are straightforward and highly reliable:

- Pronounce *-ed* as /əd/ after a stem ending in *t* or *d*. Pronounce all other cases of *-ed* as /t/ or /d/, e.g., *patented*, *decided*, *preached*, *sneezed*.
- Pronounce *-s* and *-'s* as /əz/ after a stem ending in a clue letter (e.g., *ce*, *ge*, *s/se*, *z/ze*, *ch/che*, *sh/she*, *x/xe*). Pronounce all other cases of *-s* and *-'s* as /s/ or /z/, e.g., *changes*, *Chase's*, *grips*, *Pat's*, *homes*, *Nora's*.

It is good to remember that these patterns do not exist in isolation. As illustrated in the preceding section on predicting compression, the *-ed*, *-s*, and *-'s* patterns predict segments that are subject to palatal assimilation (*You included_yourself*, *It taxes_your brain*) and linking (*closed_it*, *years_ago*). Furthermore, *-ed* has forms that can undergo cluster trimming (*mov'(ed) quickly*).

Predicting variability

With no single standard for educated pronunciation, English speakers are not uniform in the phonemes they use for some words. Depending on their dialect, we hear educated speakers say *class* with /ɑ/ and /æ/, *garage* with /ʤ/ and /ʒ/, *roof* with /uw/ and /ʊ/, and *where* with /hw/ and /w/. The unspoken protocol among educated speakers is the Golden Rule: give others the same latitude to use their own variants as we expect others to give us. We find it entirely justified to extend the same accommodation to learners of English. That is, we do not insist that learners settle on a single pronunciation for *class*, *garage*, *roof*, or *where*. We allow, even encourage, them to select the educated variant that they find easiest to pronounce, even if it is different from the teacher's usage.

To implement this policy toward variability, ESL/EFL teachers need to know where educated speakers use different pronunciations for the same word. This is not hugely challenging for teachers because variability is largely regular, being governed by environment. The same spelling-based patterns that describe consonants, word stress, vowels, and compression also describe the variability in each area.

For example, among consonants, by knowing that wh = /hw/ or /w/, we can tell that words like *when*, *where*, and *why* will be pronounced acceptably as /hw/ or /w/. Speakers of British English prefer major stress on the second syllable of two-syllable *-ate* verbs, whereas speakers of American English prefer it on the first syllable, e.g. *rotáte* versus *rótate*. The majority of phonological variability is found among vowels. Vowel prediction patterns reflect that variation. For instance, aú = /ɑ/ or /ɔ/ tells us that educated speakers may pronounce *daughter* and *cause* in two different ways. We regularly hear phonological variability among NATRL devices because all compression devices except vowel reduction are inherently variable. Educated speakers are not obliged, for example, to use palatal assimilation when saying *I miss you*, nor must they drop the middle vowel of *company* to make it *comp'ny*. The presence of phonological variability is identified in all the prediction patterns presented in Dickerson (2004).

The practical ramifications of having access to phonological variability are that we can implement a policy of tolerance toward educated variability in our pedagogical materials, our teaching, and our correction. In NATRL areas where variability is the norm, we inform learners about their range of options, when each option is appropriate to use, and encourage learners to use them when they can. In other areas where there are variable and nonvariable words, the approach is to teach the target segment or stress pattern using nonvariable words and to leave variable words to be treated as a separate group in which we monitor learners' pronunciation to see that it is within the range of acceptable variation. By being able to identify exercise items in which variability exists, we can also more easily inform our students about where different variants are available to them. Finally, when we monitor and correct our students' production,

we know where to offer leeway for alternate pronunciations among variable words and where to insist on a particular target among nonvariable words (Dickerson 1977).

Conclusion

The value to learners of knowing how to use standard orthography to predict the sounds of spoken English is enormous. That is why the loss to learners is equally great when we do not take the time to show them how to use this valuable resource for their benefit (Hill and Beebe 1980; Kreidler 1972).

One source of reticence on the part of teachers may be their deep-seated distrust of our spelling system, perpetuated by a drumbeat of largely misplaced criticism about how poorly it represents spoken words. To help teachers get beyond this barrier and come to appreciate the wealth of guidance that spelling can provide, this chapter opened with a direct challenge to conventional thinking about how English orthography actually works despite its admitted infelicities.

The message is simple: English spelling does not work the way people think it should, namely, based on straightforward spelling-to-sound correlations. Instead, first and foremost, words are spelled to communicate meaning to the eye of the reader. It does this directly by using similar spellings for related words, e.g., *write*, *writ*, *writing*, *written*, *wrote*, and different spellings for unrelated words, e.g., *write*, *right*, *wright*, *rite*. This is the operation of the meaning-first principle. Only secondarily and indirectly does spelling signal a pronunciation to the mouth of the reader. The sound-second principle works only by means of rules that readers bring to the task, e.g., wr = /r/; igh = /ay/.

With the inner workings of English orthography exposed, this chapter proceeded to show that English spelling is much more consistent than expected and much more valuable to learners who want to improve their spoken English than has been supposed. The learner rules we have presented are evidence of these claims.

We close with two cautions to the teacher. Firstly, rule-based prediction of sound from spelling is not an approach about which the teacher must decide: Do I buy in or opt out? It is not a monolithic system but a collection of many useful subsystems. Each subsystem can stand largely on its own and be integrated into pronunciation instruction without the teacher having to commit to any other parts of the system.

Secondly, it is important that teachers be realistic in what can be achieved. Since learners must access sound information indirectly through rules, prediction skills cannot be developed quickly. This means that teachers must be strategic in selecting the prediction subsystems they teach so that learners will have time to accumulate prediction skills in areas of their greatest need.

The richness of our orthography and the clarity of learner rules should be sources of encouragement to teachers and learners alike. Any part of the prediction system that we can offer our students makes their prospects brighter because that

part will become an internal resource for self-monitoring, self-correction, and self-practice as they continue to improve their oral accuracy and fluency after formal instruction ends.

REFERENCES

Bolinger, D. 1986. *Intonation and Its Parts*, Stanford, CA: Stanford University Press.

Celce-Murcia, M., Brinton, D., and Goodwin, J. 2010. *Teaching Pronunciation: A Course Book and Reference Guide*, 2nd edition, Cambridge: Cambridge University Press.

Chomsky, N. and Halle, M. 1968. *The Sound Pattern of English*, New York: Harper & Row.

Dickerson, W. 1977. The problem of dialects in teaching pronunciation. *English Teaching Forum* 15(2): 18–21.

Dickerson, W. 1980. Bisyllabic laxing rule: vowel prediction in linguistics and language learning. *Language Learning* I: 317–329. (Erratum, *Language Learning* 31: 283, 1981.)

Dickerson, W. 1990. Morphology via orthography: a visual approach to oral decisions. *Applied Linguistics* 11: 238–252.

Dickerson, W. 1994. How English orthography *really* works. In: *Current Trends: New Approaches to EFL Teaching*, C. Dirksen (ed.), 89–95, English Language Institute/China.

Dickerson, W. 2000. Covert rehearsal as a bridge to accurate fluency. Paper read at the 34th Annual TESOL Convention and Exposition, March 18, 2000, Vancouver, BC, Canada.

Dickerson, W. 2004. *Stress in the Speech Stream: The Rhythm of Spoken English*, Urbana, IL: University of Illinois Press [CD-ROM]. Available from: dickrson@illinois.edu (Original work published 1989).

Dickerson, W. 2006. A is for Alphabet … Roman letters and their use in written English. *Educators and Education Journal* 21: 1–21.

Dickerson, W. 2012. Prediction in teaching pronunciation. In: *The Encyclopedia of Applied Linguistics*, C.A. Chapelle (ed.), Hoboken, NJ: Blackwell Publishing Ltd. http://onlinelibrary.wiley.com/doi/10.1002/9781405198431.wbeal0950/full.

Field, J. 2005. Intelligibility and the listener: the role of lexical stress. *TESOL Quarterly* 39: 399–423.

Follick, M. 1965. *The Case for Spelling Reform: With a Foreword by Sir William Mansfield Cooper*, Manchester University Press.

Gilbert, J. 2001. *Clear Speech from the Start*, Cambridge: Cambridge University Press.

Guierre, L. 1984. *Drills in English Stress-Patterns,* 4th edition, Paris: Armand Colin-Longman.

Hahn, L. and Dickerson, W. 1999. *Speechcraft: Discourse Pronunciation for Advanced Learners*, Ann Arbor, MI: University of Michigan Press.

Hill, C. and Beebe, L. 1980. Contraction and blending: the use of orthographic clues in teaching pronunciation. *TESOL Quarterly* 14: 299–323.

Kreidler, C. 1972. Teaching English spelling and pronunciation. *TESOL Quarterly* 6: 3–12.

Morley, J. 1994. A multidimensional curriculum design for speech-pronunciation instruction. In: *Pronunciation Pedagogy and Theory*, J. Morley (ed.), 64–91, Alexandria, VA: TESOL.

O'Neil, W. 1980. English orthography. In: *Standards and Dialects in English,* T. Shopen and J.M. Williams (ed.), 63–83, Cambridge, MA: Winthrop Publishers, Inc.

Otto, W. and Chester, R. 1972. Sight words for beginning readers. *The Journal of Educational Research* 65: 435–443.

Vernick, J. and Nesgoda, J. 1980. *American English Sounds and Spellings for Beginning ESL Students*, Pittsburgh, PA: University Center for International Studies.

Prator, C. (1951). *Manual of American English Pronunciation for adult foreign students*, New York: Rinehart & Company, Inc.

Prator, C. and Robinett, B. 1972. *Manual of American English Pronunciation*, New York: Holt, Rinehart and Winston.

Pyles, T. 1964. *The Origins and Development of the English Language.* New York: Harcourt, Brace & World, Inc.

Sardegna, V. 2009. Improving English stress through pronunciation learning strategies. PhD dissertation, University of Illinois at Urbana-Champaign (UMI No. 3363085).

Teschner, R. and Whitley, M.S. 2004. *Pronouncing English: A Stress-Based Approach*, Washington, DC: Georgetown University Press.

Trager, E. and Henderson, S. 1957. *Pronunciation Drills for Learners of English*, Washington, DC: American University Language Center.

Weinstein, N. 2001. *Whaddaya Say?* 2nd edition, New York: Addison Wesley Longman.

Woods, H. 1979. *Syllable Stress and Unstress*, Hull, Quebec: Canadian Government Publishing Centre.

28 Technology and Learning Pronunciation

REBECCA HINCKS

Introduction

The use of technology for training pronunciation has been praised for being consistent and tireless from the time of the early phonograph to the present (e.g., Clarke 1918; Engwall et al. 2004; Levis 2007). Language educators have long been called early adaptors of new technologies (Last 1989; Roby 2004). Edison's phonograph, first commercially marketed in the 1880s, was put to use for language learning purposes as early as 1893 (Léon 1962), and its successor, the gramophone, went on to be used for the provision of native-speaker pronunciation models throughout most of the twentieth century. The introduction of magnetic tape and tape-recording machines in the period following World War II allowed the development of language laboratories (Hocking 1954), where students sitting in isolated booths or carrels could listen to speaker models and record their own pronunciations. Digital technologies were used for language learning starting in the 1960s with the PLATO project, where by the end of the 1970s over 50 000 hours of language training in a dozen languages had been developed (Hart 1995). These early computer-assisted language-learning projects had limited capacity to provide oral and aural training, but when personal computers began to be equipped with audio input and output capacity in the 1990s, learners became able to record and listen to their own versions of modeled pronunciations. Though the first pronunciation software used the computer as little more than a record-and-playback device, steps had been taken toward creating systems to provide automatic feedback on pronunciation quality.

Good computer-assisted pronunciation training (CAPT) systems can allow training to be individualized and maximized. Specific exercises can be selected to meet a learner's particular problems. The opportunity to practise is not limited to the time a teacher is available, and since a computer is infinitely patient, the time on task can be increased. However, it can be difficult to fit CAPT into a theoretical framework of how language learning best takes place. In general, computers lend

The Handbook of English Pronunciation, First Edition. Edited by Marnie Reed and John M. Levis.

themselves most naturally to the kind of training advocated by audiolingual theorists: drills, repetitions, and mimicry. The theories of the communicative approach to language learning are harder to put into practice. Further advances in artificial intelligence are necessary before computers can offer an environment that can be said to truly either "communicate" or "negotiate" with a learner, though research in creating spoken dialogue systems for language learning is underway. In future systems, mispronunciation could be part of the reason for a breakdown in communication, thus pushing students to focus on their pronunciation while providing an imitation of the "negotiation of meaning" that takes place between humans.

CAPT systems ideally use speech technology to give feedback on pronunciation. The three main components of speech technology are speech analysis, speech recognition, and speech synthesis. Speech analysis provides an acoustic analysis of the speech signal, usually in the form of a visualization of the speech waveform, spectrogram, and pitch contour. Freely available tools that provide speech analysis are, for example, Praat (Boersma and Weenink 2010) and WaveSurfer (Sjölander and Beskow 2000). An example of a WaveSurfer analysis is shown in Figure 28.1.

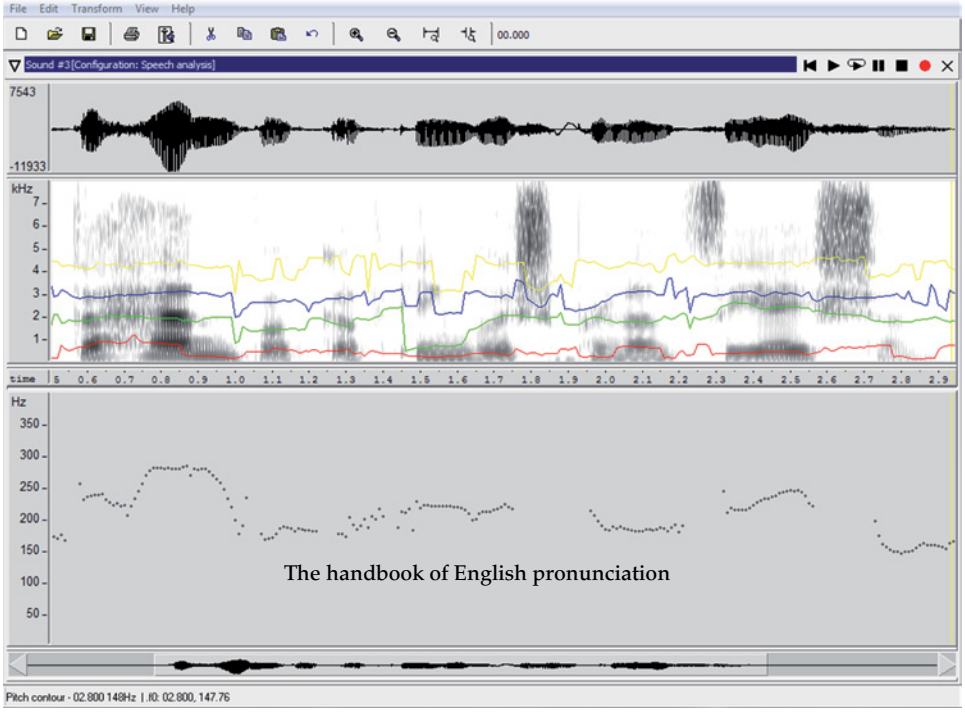


Figure 28.1 Speech analysis of the utterance "the handbook of English pronunciation", showing the speech waveform (top), spectrogram with colored tracings of the first four formants (middle), and pitch contour (bottom).

Automatic speech recognition, also known as ASR, voice recognition, or speech-to-text (STT), turns the speech signal into words (for an introduction see Ainsworth 1997). Commercial ASR systems, such as Dragon Dictate, have taken longer than was originally hoped for to reach reliable word error rates, but are now reaching a wide mass market as software packages and apps.

Computer speech is generated by speech synthesis, or TTS (text-to-speech) (for an introduction see Carlson anad Granström 1997). There are freely available synthesis programs (e.g., http://espeak.sourceforge.net/index.html) that make use of the rule-based approach known as formant synthesis; however, the most natural-sounding synthesis is produced by splicing together tiny bits of sound taken from a large database of recordings of the speech of one individual. When combined with systems for natural language understanding and processing, ASR and TTS are used for dialogue systems, now being used in some speech training applications.

The remainder of this chapter is organized not by the technologies themselves but by the kind of feedback on learner production that they can provide. We will look first at technologies for capturing and modeling pronunciation. The next two sections discuss technologies for feedback on suprasegmental and segmental production respectively. We then look at how speech technologies can assess pronunciation, and finally at how technologies can provide conversational practice.

Technology for capturing and modeling pronunciation, with limited feedback

This section briefly considers how technologies for recording, playing back, and synthesizing speech have been used in ways that were innovative in their times.

Record and play back

The earliest language labs, using gramophones, were set up with the purpose of giving students opportunities to individually mimic recorded models, though from the outset questions were raised about the benefits of imitation without feedback (Roby 2004). The postwar language lab, using magnetic tape, offered students the ability to not only listen to models but also to record their own speech and listen to it. After their heyday in the 1960s, language labs fell rather abruptly out of fashion in the 1970s (Roby 2004). Research had been unable to show definitively that lab use had a positive effect on language proficiency, and the installations were described as expensive "electronic graveyards" (Turner 1969). The disillusionment with language labs was part of the transition from audio-lingual or behaviorist theories of language learning to constructivist models.

The successor to the language lab was the computer lab. CALL software by the 1990s was using "multimedia", that is, audio, video, and graphics, to contribute to the learning process. When it came to pronunciation, however, most of the

commercial software that claimed to teach pronunciation used the computer as no more than a recording device as late as the year 2000 (Hincks 2005a). With these programs, students were still expected to self-diagnose their pronunciation weaknesses by listening to recordings of their own speech collected by the software.

As "digital natives" have filled classrooms in the twenty-first century, we have seen innovative use of audio technology with the ultimate purpose of improving pronunciation. The English language, with its breadth of global, regional, and social pronunciation models, is especially well-suited to using technology to provide access to recordings of varieties that are remote from an individual setting. The Internet encompasses an enormous wealth of models of pronunciations for students in a lingua franca environment who would benefit from understanding many varieties of English. The web is also used to disseminate phonetic knowledge (e.g., by the University of Iowa's phonetics department, http://www.uiowa.edu/~acadtech/phonetics/). Language students can also use the Internet to distribute oral texts – podcasts – and in that way improve pronunciation (e.g., Lord 2008). The web is also the distribution medium for low-cost apps for pronunciation training, with or without feedback.

Feedback on the perception of pronunciation can be given without the use of speech processing. While it is a difficult task for a computer to provide good feedback on a student's production of pronunciation, it is a simple one to give feedback on a student's perception of pronunciation. Computer programs can successfully be used to practise the perception of minimal pairs and lexical stress placement (Wik, Hincks, and Hirschberg 2009) or streams of rapid spontaneous speech (Cauldwell 1996). The program contributes feedback in the form of telling the student whether a response was correct or not; this is limited feedback, but is a first step to achieving good pronunciation.

Generating speech with synthesis

One as yet relatively unexplored means of providing pronunciation models is through the use of speech synthesis. The greatest research challenge at present is to improve the naturalness of synthesis, which can be done by finding better ways to choose what prosodic contour should best be applied to an utterance. Because synthesis can often sound quite artificial, developers have been wary of using it as a teaching model, preferring recordings of natural voices. Work has been done to develop a methodology for benchmarking synthesis so that it can be more reliably used in CALL applications (Handley 2009; Handley and Hamel 2005).

The advantage of using synthesis in any speech application is that it eliminates the reliance on the existence of pre-recorded utterances, which need to be planned in advance in order to maintain a consistent speaker voice. Any utterance can be generated at any time. Text-to-speech synthesis as a widely available learning tool would empower learners to generate the pronunciation of utterances in the absence of authoritative speakers of the language. Recent advances made in the naturalness of commercially available synthesis systems have inspired their use as reading models in situations where teachers may either not have satisfactory

pronunciation or time to record large quantities of text. Students can thereby listen to a text as they read it, in that way doubling the channels of linguistic input. Speech synthesis could also potentially be used to disseminate new models of English. Jenkins (2000), for example, took the bold step of proposing a new standard for spoken International English. Since the Jenkins variety of English would be an artificial construction, it has no native speakers, and could be modeled and disseminated by the use of speech synthesis.

Speech synthesis can be manipulated with a level of control that cannot always be achieved with natural speech, and therefore it is often used to test the perception of speech sounds. The goal of this type of perception research has been an understanding of the relevant acoustic properties of speech sounds and how humans perceive them. For language learners, it is generally believed that perceiving second language sound contrasts is a prerequisite to being able to produce them, and it has been shown that they need to be exposed to a variety of voices in order to be able to generalize trained perception of L2 sound contrasts to new stimuli (Lively, Logan, and Pisoni 1994). Though these researchers achieved their results with recordings of natural speech, synthesis is an alternative for producing stimuli for the purpose of teaching the perception of L2 sounds. Wang and Munro (2004) successfully used synthetic stimuli to teach Mandarin and Cantonese learners distinctions in English vowel quality. With the goal of teaching students to focus less on vowel duration and more on vowel quality, they used formant synthesis to create stimuli with six different vowel durations. For example, the words heed and hid were each synthesized with different vowel durations ranging between 125 to 250 ms. The students thereby learned to listen to the differences in quality rather than length to distinguish between /i/ and /I/. Long-term improved perception of the contrasts in comparison with a control group was achieved.

One potential for speech synthesis in CAPT applications lies in its ability to be freely integrated with visual models of the face, mouth, and vocal tract (Engwall 2012). The visual component is an important part of spoken language understanding (Grant and Greenberg 2001) and is clearly essential when it comes to pronunciation instruction (Elliot 1995). Traditional CAPT systems use videos of human faces or animations to demonstrate correct articulation. Future systems will be able to reveal what the articulation should look like inside the oral cavity as well as on the outside of the mouth. This will provide important information about articulation, for example, for tongue placement.

Technology for suprasegmental feedback

Speech analysis software has been used to give visual feedback on intonation since the 1960s. The basic principle is that the pitch contour and sound waveform (see Figure 28.1) of a student utterance are displayed alongside those of a model utterance. Together with a teacher, or on his or her own, the student examines the differences in the visualizations of the two utterances, with the goal of achieving a better match in terms of pitch, duration, and intensity.

Pitch visualization lends itself most naturally to training short utterances that rely on pitch movement to distinguish meaning. For English, this could include, for example, polar-question intonation, pitch movement on key words, and minimal pairs distinguished by lexical stress placement. Longer discourse is harder to display and interpret at a fine-grained level, but can be useful to illustrate, for example, raised pitch with the introduction of a new topic, falling pitch across the length of an utterance, and tone choice (Levis and Pickering 2004). The speech waveform, when properly interpreted, reveals information about relative syllable length so that learners can observe durational differences between stressed and unstressed syllables.

Studies have shown that presenting visual displays of pitch contours improves both perception and production of target language intonation. Groundbreaking work was done in the Netherlands by de Bot and Mailfert (1982) who showed that even limited training with audiovisual feedback of prosody was more beneficial than audio feedback alone. A similar line of investigation was later carried out by Molholt (1988) on Chinese-speaking learners of English and by Oster (1998) on immigrants to Sweden. Hardison (2004) expanded this type of work to show that audiovisually trained learners of French not only improved their prosody but also their segmental accuracy.

These successful studies, however, were conducted largely in situations where there was a teacher available for guidance and interpretation. Because most language learners have little knowledge of acoustics, they need help to understand pitch displays. Pitch contours consist of not-always-intuitive broken lines, where the unvoiced segments of speech, which lack fundamental frequency, cause gaps that can be disconcerting to a learner. Furthermore, if a student and a model speaker have very different natural voice ranges, it can be hard to see the relationship between two pitch curves, and they may not be displayed with proper alignment with each other. Other problems can be caused by the algorithms for pitch extraction from the speech signal, which do not always work perfectly. Miscalculation of the fundamental frequency can lead to sudden discrepancies of an octave or more, so that the pitch contour suddenly seems to disappear. For best results, when signal analysis software is used to show intonation, it should be calibrated to respond to the vocal range of an individual speaker.

Speech analysis for visualizing intonation has not been widely used in language classrooms. Initially, teachers were put off by the high cost of signal analysis software such as VisiPitch (Chun 1998) or SpeechViewer (Oster 1998). The systems freely available, such as Praat (Boersma 2001) and WaveSurfer (Sjölander and Beskow 2000), were for some time well known in the speech research community but relatively unknown in the language-learning community. Some researchers/teachers have pointed out that the necessity of using utterances with many voiced (as opposed to unvoiced) segments presents an obstacle to its use (Anderson-Hsieh 1992; Chun 1998).

Automatically comparing two pitch contours for the purposes of supplying CAPT feedback is not a simple task. Research efforts are under way to apply pattern recognition and matching techniques to evaluate learner placement of

lexical stress in English (Hönig et al. 2010), and there are commercial CAPT packages that incorporate a signal analysis element, though it is unclear in what way they use the intonation information in their feedback. It is unlikely that longer utterances could ever be meaningfully compared automatically. In an effort to find new ways of automatically using pitch information, Hincks (2005b) suggested that only the pitch data, rather than the visualized contours, be used as feedback. Pitch variation correlates with perceptions of speaker liveliness, which is important in public speaking and can be difficult to achieve when speaking in a second language. In a later study, Hincks and Edlund (2009) gave real-time feedback on pitch variation as Chinese learners of English practised oral presentations. The feedback was successful in teaching the students to speak with more liveliness.

One technique that is theoretically appealing is resynthesis of a student's own production (Bannert and Hyltenstam 1981; De Meo et al. 2013; Sundström 1998). In resynthesis, the pitch and duration parameters of a native speaker are applied to an utterance made by a language learner. Providing the original utterance had acceptable segmental quality, the result is that the student is able to hear his or her own voice sounding much more like a native speaker. Listening to one's own resynthesized utterance should lower some of the psychological barriers to adapting the intonation patterns of the target language. Felps, Bortfeld, and Gutierrez-Osuna (2009) applied the technique to a corpus of learner utterances, and evaluated the perception of the resynthesized versions. Their resynthesis was shown to reduce the perception of foreign accentedness while maintaining the voice quality properties of the foreign speaker. De Meo et al. (2013) found that self-imitation was more effective than imitation of a standard model in training Chinese speakers to achieve Italian prosodic patterns.

Technology for giving feedback at the segmental level

Since the mid-1990s, automatic speech recognition has been used in CAPT systems. ASR holds the tantalizing promise of enabling a communicative, feedback-providing framework for CALL, by letting learners "converse" with a computer in a spoken dialogue system. ASR technology has improved greatly in recent years, and reached a type of mass market with the growing use of voice applications such as Apple's Siri in mobile devices. However, significant advances in natural language processing and computational power are necessary before even native speakers can converse with a computer about anything beyond the constraints of limited domains. These challenges are multiplied for the prospect of accented users using speech recognition, since their pronunciations cannot be represented in a general language database without diluting the precision of the recognition (Egan and LaRocca 2000). For the time being, ASR can be used in CAPT systems to give automatic feedback on the quality of phoneme production.

The basis of ASR technology is the probabilistic comparison between the signals received by the system and what is known about the phonemes of a language as represented in a database containing recordings of hundreds of native speakers of

the language (Ainsworth 1997). Because of ASR's mathematical basis, numerical scores can be derived representing the deviation between a signal and an acoustic model of the phoneme it is hypothesized to have initiated from. These scores have the potential to then be given to the learner as a type of feedback measuring a quantifiable distance from a target phoneme. However, it is not possible with current technology to say in what way the signal has deviated from the model, and this means that feedback is not corrective or constructive, but merely a sort of evaluation of the signal. Neri et al. (2002) raised the issue of whether the use of ASR in CAPT systems was driven by technology or by pedagogy, and proposed guidelines for the successful systems for teaching Dutch later developed by their research group at Radboud University in the Netherlands.

In a typical ASR-based CAPT system, a prompt will be given to a student, who can then choose a response from a limited set. One way to do this is to present a number of alternatives that the student basically can read up, and another is to design questions that can be answered only in very limited ways. Even if the student is heavily accented, the ASR system can still have a good chance of recognizing the answer if the choices are limited. Once the student response is recognized, it is aligned segment by segment with the model version in the system, and compared to find what sounds most deviate from the model sounds the ASR is based on. A well-designed CAPT system (Cucchiarini, Neri, and Strik 2009) has predetermined pedagogical priorities as to what sounds are most important to give feedback on, based on their functional load within the language. Another issue the system needs to take into consideration is what is known as the error threshold, which refers to the degree of certainty that a student has produced a correct or incorrect pronunciation. Systems can be tuned as to whether they should lean in favor of falsely accepting incorrect responses or falsely rejecting correct ones; it is probably better to do the former rather than the latter. Finally, it is wise to limit the amount of corrective feedback to avoid overwhelming the student.

Research has shown that carefully designed ASR-based training can produce positive results in teaching learners how to produce targeted sounds, such as the /x/ sound in Dutch (Cucchiarini, Neri, and Strik 2009). However, the kinds of studies that are possible to do in most real-world contexts, with many sometimes uncontrollable variables such as student engagement or the time on task, have been unable to show pronunciation development that has expanded from the improvement of a limited number of sounds to any kind of better pronunciation from a more holistic perspective (Cucchiarini, Neri, and Strik 2009; Hincks 2003). One reason for this could be the fact that speech recognition systems at present are poor at handling information contained in the speaker's prosody. In order to recognize the words of an utterance, the recognizer must ignore the variations of pitch, tempo, and intensity that naturally appear in utterances by different speakers and even within an individual speaker's various productions. This means that ASR can give feedback at the segmental level, but not on the suprasegmental level (with the exception of speaking rate, which will be handled below). Unfortunately for CALL developers who want to use ASR, these prosodic features are sometimes those that need the most practice from language learners (Anderson-Hsieh,

Johnson, and Koehler 1992). Another possible reason for the relatively disappointing performance of ASR-based CAPT could be that critics of audiolingual language training were right: drilling and mimicry are not the best way to learn pronunciation.

ASR-based CAPT would be improved if the feedback could give precise instruction as to how a sound could be better articulated. One way to do that would be by working with specific L1-L2 pairs. It is known, for example, that German learners of English have a problem with devoicing in word-final position, so if the second consonant in the word "rise" produced in a CAPT system receives a low score, then the learner could be automatically prompted to voice the sound. Creating systems like this might be possible for pairs of the world's major languages, but it is a very expensive process. The ASR speech database would need to consist of carefully annotated German-accented English, mixed with native-accented English. Since the global market for learning English is so valuable, this might be a worthwhile process for a commercial operation, but what about Somali-native learners of Swedish? Such specific systems will of course never exist, and without them the ASR will only be able to give scores on pronunciation without feedback on how to improve articulation.

In addition to the question of better feedback, there are a number of other issues regarding the ASR speech models used in CAPT systems. Many language learners are children, but their speech is not suitable for recognition in systems based on recordings of adult speech, and special databases and programs need to be created for them (Elenius and Blomberg 2005; Gerosa and Giuliani 2004). With the creation of such a system, Neri et al. (2008) were able to show that Italian children studying English learned the pronunciation of new vocabulary as well from a computer system as from a teacher. Ideally, users should model their utterances on those of speakers of the same sex. Work on allowing users to pick their own model speaker was done by Probst, Ke, and Eskenazi (2002); unfortunately, users were not very successful in choosing models that were appropriate for their voices. Another issue perhaps specific to English and a few other languages is the fact that there is more than one standard teaching model of English. It would be discouraging for a student who has been taught British English to receive negative feedback in a CAPT system that used underlying American English models.

The most widely known application for ASR is the spoken dialogue systems with which we can, for example, order tickets automatically, but an application available on the consumer market is for computer-based dictation. Dictation systems were previously speaker-dependent, that is, trained to recognize the speech of one individual. Recent breakthroughs in ASR technology have allowed the development of dictation systems that are speaker-independent. That is, they are able to recognize any speaker's voice. A few researchers have been inspired to test dictation systems on language learners, as a way of assessing pronunciation. Coniam (1999) looked at the ability with which foreign-accented speakers of English could use a commercially available speaker-dependent dictation program. Predictably, the software was significantly worse at recognizing foreign-accented speech than native speech. Derwing, Munro, and Carbonaro (2000) compared

ASR recognition scores with human intelligibility scores derived by transcribing recorded utterances. Like Coniam, they found that proficient non-native speech was recognized much less accurately than native speech; moreover, they found a discrepancy between errors perceived by humans and the misrecognitions of the dictation software. The problems the dictation systems encountered did not correspond to a human-like pattern as evidenced by human intelligibility scores. It is important to remember, however, that dictation software has not been designed with CAPT applications in mind. ASR for non-native speech needs to be adapted so that the underlying phonetic models encompass a wider variety of possible productions.

Because of the inherent limitations in the way standard ASR can be used for CAPT, researchers are testing other ways of using speech processing for feedback at the segmental level, though these methods are not as automated. Researchers have let students practise single words or phrases with visual feedback in the form of a spectrogram (provided by speech analysis software) and in the presence of a teacher for guidance and interpretation (Pearson, Pickering, and Da Silva 2011; Ruellot 2011). Cutting-edge work by Engwall and co-authors (Engwall 2012; Engwall and Balter 2007; Kjellström and Engwall 2009) has looked at what sorts of supplemental information can provide clues to the causes of deviant pronunciation. Their idea is to make use of features in the acoustic signal that indicate articulatory information, e.g., place or manner, and furthermore combine the acoustic information with visual information from a speaker's mouth and face. This information can then be used to create feedback that gives instruction about better articulation, instead of the mere classification into acceptable or not-acceptable phonemes that can be given by traditional techniques.

Technology for evaluating pronunciation

An obstacle in testing pronunciation is determining a practical method for evaluation. Human judgment is not only time consuming and expensive, but it sometimes can be difficult for raters to be consistent and to agree with others. An appealing alternative would be to let ASR provide an objective measure for a pronunciation test. Since ASR is better at quantifying deviation from a norm than providing corrective feedback, pronunciation evaluation is in fact its most natural application in the language learning field. ASR can also be used to determine whether a student has given the correct answer to a simple question, such as "What is the opposite of 'complex'?" Questions like this can be used to assess a student's vocabulary and thereby language proficiency. Furthermore, ASR can be easily used to measure the speed at which a learner speaks, a type of fluency measure. Rate of speech has been shown to correlate with speaker proficiency (Cucchiarini, Strik, and Boves 2002; Hincks 2010; Kormos and Dénes 2004). Thus, the best prosodic application of ASR is in the assessment of temporal measures.

With the aim of creating an automatic pronunciation test for spoken Dutch, Cucchiarini, Strik, and Boves (2000) devised an extensive study that looked at the

correlations between different aspects of human ratings of accented Dutch and machine scores provided by ASR. They found a high correlation between human ratings and machine-generated temporal measures such as rate of speech and total duration. In other words, speakers judged highly by the raters were also the faster speakers. However, the ASR in this system did a poor job of assessing segmental quality, which was the aspect of speech that the human raters found to be most important when rating accentedness. There was thus a mismatch between what humans associated with good speech and what computers rated as good speech. However, the ASR was still able to discern the better speakers; it just used another way of finding them than the humans did.

The commercially successful Versant (formerly PhonePass) test (Bernstein et al. 2000; Bernstein, Van Moere, and Cheng 2010) uses speech recognition to assess the correctness of student responses, and also gives scores in pronunciation and fluency. Comparisons of the results given by the test and those obtained by human-rated measures of oral proficiency show that there is as much correlation between its scores and averaged human ratings as there is between one human rater and another (Bernstein et al. 2001). An examination of the PhonePass scores of a group of students was published in Hincks (2001). That paper found a relationship between the speed at which students read test sentences and the scores they received, and discussed the risks inherent in assessing short, nonphonetically balanced samples of speech.

Technology for practising speaking skills

The market for CALL systems for English is enormous, especially in Asia. It is estimated, for example, that nearly 2% of the Korean GNP is spent on learning English (Pellom 2012). Companies that produce products for these markets are aware of the serious limitations in the ability of speech processing techniques to provide accurate formative feedback that can achieve measurable improvement in pronunciation. Some of them have therefore shifted the focus of their products from CAPT to the more general "practice in speaking", with, for example, accompanying social and cultural training (Johnson 2012). Training delivered by the Internet provides an opportunity for human teachers to come in to give feedback on pronunciation after a student has practised a dialogue, a strategy adopted by a major American company (Pellom 2012) in its high-end systems.

One dream of CALL developers is the use of unconstrained dialogue systems for language speaking practice. A dialogue system combines speech recognition, natural language understanding, and speech synthesis to enable a person to communicate with a computer and complete a task. Developers are working on "embodied conversational agents" that can act as both language tutors and conversational partners. A number of projects include a gaming element, where, for example, a learner must bargain for a product in a flea-market environment (Wik and Hjalmarsson 2009), quickly provide a translation of a word (Seneff 2007) or exhibit culturally sensitive behavior (Johnson, Vilhjalmsson, and Marsella 2005). Games are believed to stimulate engagement and learning in a nonthreatening environment.

Conclusion

There is an enormous need for CAPT, a need expressed in a number of review articles in recent years (Eskenazi 2009; Levis 2007; O'Brien 2006). However, really effective automated feedback remains an elusive goal; in the words of one developer, it is a problem that is not too big to run away from (Johnson 2012). Until the research challenges for automation are solved, teachers are encouraged to work with students individually or in small groups, using proven methods to raise pronunciation awareness. The studies that have shown the most convincing benefits to learners (Hardison 2004; Pearson, Pickering, and Da Silva 2011) have used speech analysis software such as freely available Praat and WaveSurfer, and have not eliminated the presence of the teacher. The field of pronunciation training has a long tradition of embracing new technologies, and speech visualization is one of them.

REFERENCES

Ainsworth, W.A. 1997. Some approaches to automatic speech recognition. In: *The Handbook of Phonetic Sciences*, W.J. Hardcastle and J. Laver (eds.), Oxford, Blackwell.

Anderson-Hsieh, J. 1992. Using electronic visual feedback to teach suprasegmentals. *System* 20(1): 51–62. doi: 10.1016/0346-251X(92)90007-P.

Anderson-Hsieh, J., Johnson, R., and Koehler, K. 1992. The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody and syllable structure. *Language Learning* 42(4): 529–555. doi: 10.1111/j.1467-1770.1992.tb01043.x.

Bannert, R. and Hyltenstam, K. 1981. Swedish immigrants' communication: problems of understanding and being understood. *Working Papers Lund University Department of Linguistics* 21: 17–28.

Bernstein, J., Van Moere, A., and Cheng, J. 2010. Validating automated speaking tests. *Language Testing* 27(3): 355–377.

Bernstein, J., Jong, J.D., Pisoni, D., and Townshend, B. 2000. Two experiments on automatic scoring of spoken language proficiency. Paper presented at the Integration of Speech Technology in Language Learning, Dundee, Scotland.

Boersma, P. 2001. PRAAT: a system for doing phonetics by computer. *Glot International* 5: 341–345.

Boersma, P. and Weenink, D. 2010. Praat: doing phonetics by computer (Version 5.1.41) [computer program]. Retrieved July 22, 2010 from: http://www.fon.hum.uva.nl/praat/.

Carlson, R. and Granström, B. 1997. Speech synthesis. In: *The Handbook of Phonetic Sciences*, W.J. Hardcastle and J. Laver (eds.), Oxford, Blackwell.

Cauldwell, R.T. 1996. Direct encounters with fast speech on CD-audio to teach listening. *System* 24(4): 521–528. doi: 10.1016/s0346-251x(96)00046-2.

Chun, D. 1998. Signal analysis software for teaching discourse intonation. *Language Learning and Technology* 2(1): 61–77.

Clarke, C.C. 1918. The phonograph in modern language teaching. *Modern Language Journal* 3: 116–122.

Coniam, D. 1999. Voice recognition software accuracy with second language

speakers of English. *System* 27(1): 49–64. doi: 10.1016/S0346-251X(98)00049-9.

Cucchiarini, C., Neri, A., and Strik, H. 2009. Oral proficiency training in Dutch L2: the contribution of ASR-based corrective feedback. *Speech Communication* 51(10): 853–863. doi: 10.1016/j.specom.2009.03.003.

Cucchiarini, C., Strik, H., and Boves, L. 2000. Different aspects of expert pronunciation quality ratings and their relation to scores produced by speech recognition algorithms. *Speech Communication* 30: 109–119. doi: 10.1016/S0167-6393(99)00040-0.

Cucchiarini, C., Strik, H., and Boves, L. 2002. Quantitative assessment of second language learners' fluency: comparison between read and spontaneous speech. *Journal of the Acoustical Society of America* 111(6): 2862–2873.

De Bot, K., and Mailfert, K. 1982. The teaching of intonation: fundamental research and classroom applications. *TESOL Quarterly* 16(1): 71–77.

De Meo, A., Vitale, M., Pettorino, M., Cutugno, F., and Origlia, A. 2013. Imitation/self-imitation in computer-assisted prosody training for Chinese learners of L2 Italian. Paper presented at the 4th Pronunciation in Second Language Learning and Teaching Conference, Ames, Iowa.

Derwing, T., Munro, M., and Carbonaro, M. (2000). Does popular speech recognition software work with ESL speech? [Brief Report]. *TESOL Quarterly* 34(3): 592–602.

Egan, K. and LaRocca, S. 2000. Speech recognition in language learning: a must. Paper presented at the Integration of Speech Technology in Language Learning, Dundee, Scotland.

Elenius, D. and Blomberg, M. 2005. Adaptation and normalization experiments in speech recognition for 4- to 8-year old children. Paper presented at the Interspeech 2005, Lisbon.

Elliot, R. 1995. Foreign language phonology: field independence, attitude, and the success of formal instruction in Spanish pronunciation. *Modern Language Journal* 79(IV): 530–542.

Engwall, O. 2012. Analysis of and feedback on phonetic features in pronunciation training with a virtual teacher. *Computer Assisted Language Learning* 25(1): 37–64. doi: 10.1080/09588221.2011.582845.

Engwall, O. and Balter, O. 2007. Pronunciation feedback from real and virtual language teachers. *Computer Assisted Language Learning* 20(3): 235–262. doi: 10.1080/09588220701489507.

Engwall, O., Wik, P., Beskow, J., and Granström, B. 2004. Design strategies for a virtual language tutor. Paper presented at the International Conference on Spoken Language Processing, Jeju Island, Korea, October 4–8, 2004.

Eskenazi, M. 2009. An overview of spoken language technology for education. *Speech Communication* 51(10): 832–844. doi: 10.1016/j.specom.2009.04.005.

Felps, D., Bortfeld, H., and Gutierrez-Osuna, R. 2009. Foreign accent conversion in computer assisted pronunciation training. *Speech Communication* 51(10): 920–932. doi: 10.1016/j.specom.2008.11.004.

Gerosa, M. and Giuliani, D. 2004. Preliminary investigations in automatic recognition of English sentences uttered by Italian children. Paper presented at the Integrating Speech Technology in Language Learning, Venice, Italy.

Grant, K. and Greenberg, S. 2001. Speech intelligibility derived from asynchronous processing of auditory-visual information. Paper presented at the Audio-visual Speech Processing Workshop 2001.

Handley, Z. 2009. Is text-to-speech synthesis ready for use in computer-assisted language learning? *Speech Communication* 51(10): 906–919. doi:10.1016/j.specom.2008.12.004.

Handley, Z. and Hamel, M.-J. 2005. Establishing a methodology for benchmarking speech synthesis for

computer-assisted language learning (CALL). *Language Learning and Technology* 9(3): 99–119.
Hardison, D. 2004. Generalization of computer-assisted prosody training: quantitative and qualitative findings. *Language Learning and Technology* 8(1): 34–52.
Hart, R.S. 1995. The Illinois PLATO foreign languages project. *CALICO Journal* 12(4): 15–37.
Hincks, R. 2001. Using speech recognition to evaluate skills in spoken English. *Proceedings of Fonetik* 2001, 49: 58–61.
Hincks, R. 2003. Speech technologies for pronunciation feedback and evaluation. *ReCALL* 15(1): 3–20.
Hincks, R. 2005a. Computer support for learners of spoken English. Doctoral thesis, Royal Institute of Technology, Stockholm, Sweden.
Hincks, R. 2005b. Measures and perceptions of liveliness in student oral presentation speech: a proposal for an automatic feedback mechanism. *System* 33(4): 575–591. doi: 10.1016/j.system.2005.04.002.
Hincks, R. 2010. Speaking rate and information content in English lingua franca oral presentations. *English for Specific Purposes* 29(10): 4–18. doi: 10.1016/j.esp.2009.05.004.
Hincks, R. and Edlund, J. 2009. Promoting increased pitch variation in oral presentations with transient visual feedback. *Language Learning and Technology* 13(3): 32–50.
Hocking, E. 1954. *Language Laboratory and Language Learning*, Washington, DC: Division of Audiovisual Instruction, National Education Association.
Hönig, F., Batliner, A., Weilhammer, K., and Noth, E. 2010. Automatic assessment of non-native prosody for English as L2. Paper presented at the Speech Prosody Chicago, Illinois.
Jenkins, J. 2000. *The Phonology of English as an International Language: New Models, New Norms, New Goals*, Oxford: Oxford University Press.
Johnson, L. 2012. Error detection for teaching communicative competence. Paper presented at The International Symposium on Automatic Detection of Errors in Pronunciation Training, Stockholm, Sweden.
Johnson, L., Vilhjalmsson, H., and Marsella, S. 2005. Serious games for language learning: How much game, how much AI? In: *Artificial Intelligence in Education*, C.-K. Looietal (ed.), Amsterdam: IOS Press.
Kjellström, H., and Engwall, O. 2009. Audiovisual-to-articulatory inversion. *Speech Communication* 51(3): 195–209. doi: 10.1016/j.specom.2008.07.005.
Kormos, J. and Dénes, M. 2004. Exploring measures and perceptions of fluency in the speech of second language learners. *System* 32: 145–164. doi: 10.1016/j.system.2004.01.001.
Last, R.W. 1989. *Artificial Intelligence Techniques in Languge Learning*, Chichester, England: Harwood.
Léon, P.R. 1962. *Laboratoire de Langues et Correction Phonétique*, Paris: Didier.
Levis, J. 2007. Computer technology in teaching and researching pronunciation. *Annual Review of Applied Linguistics* 27: 184–202.
Levis, J. and Pickering, L. 2004. Teaching intonation in discourse using speech visualization technology. *System* 32: 505–524.
Lively, S., Logan, J., and Pisoni, D. 1994. Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America* 94(3): 1242–1255.
Lord, G. 2008. Podcasting communities and second language pronunciation. *Foreign Language Annals* 41(2): 364–379.
Molholt, G. 1988. Computer-assisted instruction in pronunciation for Chinese speakers of American English. *TESOL Quarterly* 22(1): 91–111.

Neri, A., Cucchiarini, C., Strik, H., and Boves, L. 2002. The pedagogy–technology interface in computer assisted pronunciation training. *Computer-Assisted Language Learning* 15(5): 441–467. doi: 10.1076/call.15.5.441.13473.

Neri, A., Mich, O., Gerosa, M., and Giuliani, D. 2008. The effectiveness of computer assisted pronunciation training for foreign language learning by children. *Computer Assisted Language Learning* 21(5): 393–408. doi: 10.1080/09588220802447651.

O'Brien, M. 2006. Teaching pronunciation and intonation with computer technology. *Calling on CALL: From Theory and Research to New Directions in Foreign Language Teaching*, L. Ducate and N. Arnold (eds.), San Marcos, TX: Computer Assisted Language Instruction Consortium.

Oster, A.-M. 1998. Spoken L2 teaching with contrastive visual and auditory feedback. Paper presented at the International Conference on Spoken Language Processing, Sydney, Australia.

Pearson, P., Pickering, L., and Da Silva, R. 2011. The impact of computer assisted pronunciation training on the improvement of Vietnamese learner production of English syllable margins. Paper presented at the 2nd Pronunciation in Second Language Learning and Teaching Conference, Ames, Iowa.

Pellom, B. 2012. Rosetta Stone ReFLEX: toward improving English conversational fluency in Asia. Paper presented at the International Symposium on Automatic Detection of Errors in Pronunciation Training, Stockholm, Sweden.

Probst, K., Ke, Y., and Eskenazi, M. 2002. Enhancing foreign language tutors – In search of the golden speaker. *Speech Communication* 37: 161–173. doi: 10.1016/S0167-6393(01)00009-7.

Roby, W.B. 2004. Technology in the service of foreign language teaching: the case of the language laboratory. In: *Handbook of Research on Educational Communications and Technology*, 2nd edition, D. Jonassen (ed.), 523–541, Mahwah, NJ: Lawrence Erlbaum.

Ruellot, V. 2011. Computer-assisted pronunciation learning of French /u/ and /y/ at the intermediate level. Paper presented at the 2nd Pronunciation in Second Language Learning and Teaching, Ames, Iowa.

Seneff, S. 2007. *Web-based dialogue and translation games for spoken language learning.* Paper presented at the Speech and Language Technology in Education (SLaTE) 2007, Pittsburgh, PA.

Sjölander, K. and Beskow, J. 2000. WaveSurfer: an open source speech tool. Retrieved August 2013 from: http://www.speech.kth.se/wavesurfer/. Paper presented at the International Conference on Spoken Language Processing 2000, Beijing.

Sundström, A. 1998. Automatic prosody modification as a means for foreign language pronunciation training. Paper presented at the ISCA Workshop on Speech Technology in Language Learning (STILL 98), Marholmen, Sweden.

Turner, E.D. (1969). Correlation of language class and language laboratory. *ERIC Focus Reports on the Teaching of Foreign Languages*, vol. 13, New York.

Wang, X. and Munro, M. 2004. Computer-based training for learning English vowel contrasts. *System* 32: 539–552.

Wik, P., Hincks, R., and Hirschberg, J. (2009). *Responses to Ville: A virtual language tutor for Swedish* Paper presented at the Speech and Language Technology in Education Warwickshire, England

Wik, P., & Hjalmarsson, A. (2009). Embodied conversational agents in computer assisted language learning. *Speech Communication*, 51(10), 1024–1037. doi: 10.1016/j.specom.2009.05.006

Index

The Handbook of English Pronunciation, First Edition. Edited by Marnie Reed and John M. Levis.

WILEY END USER LICENSE AGREEMENT

Go to www.wiley.com/go/eula to access Wiley’s ebook EULA.