

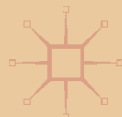
Research and Practice in Applied Linguistics
Series Editors: Christopher N. Candlin and Jonathan Crichton

English Pronunciation Teaching and Research

Contemporary Perspectives



Martha C. Pennington and
Pamela Rogerson-Revell



Research and Practice in Applied Linguistics

Series Editors

Christopher N. Candlin
Macquarie University
Sydney, NSW, Australia

Jonathan Crichton
University of South Australia
Adelaide, SA, Australia

“English Pronunciation Teaching and Research: Contemporary Perspectives breaks new ground in presenting an applied, sociolinguistic orientation to pronunciation teaching and research that is both up-to-date and comprehensive in scope. Written by two well-known pronunciation specialists, one British and one American, the book is a welcome addition to the pronunciation literature that should be on the reading lists of all language teachers and applied linguists.”

—Rodney H. Jones, *University of Reading, UK*

“This interesting and informative book makes a valuable contribution by connecting research and practice while providing a comprehensive scope. This is much appreciated given the extensive amount of research in the field as well as in related areas.”

—Jose Antonio Mompean Gonzalez, *University of Murcia, Spain*

This flagship series was created and overseen by Professor Christopher N. Candlin, and continues his work by providing the essential cross-over between research in applied linguistics and its practical applications in the professions. Books in the series address the growing need for professionals concerned with language and communication issues to keep up to date with applied linguistic research relevant to their practice. Central to this agenda, the series offers students and practising professionals rapid and authoritative access to current scholarship and research on key topics in language education and professional communication more broadly, emphasising the integration and interdependence of research and practice in a useable way. The series provides books with a common structure, each book offering a clear, up-to-date and authoritative overview of key concepts, research issues and developments in the particular topic, identifying: research evidence for the main ideas and concepts competing issues and unsolved questions the range of practical applications available for professional and organisational practice that draw on such concepts and ideas a synopsis of important issues open for action and practice-based research by practitioners/students. These state-of-the-art overviews are supported by selected cases studies of the practical applications of research and 'how to' research guides and resources, all designed to extend and localise knowledge of the topic and its relevance for the reader. Throughout the books, readers are encouraged to take up issues of enquiry and research that relate to their own contexts of practice, guided by reflective and exploratory questions and examples that invite practical connections to their work. Written by leading scholars and practitioners, the books will be essential reading for MA or PhD student in Applied Linguistics, TESOL, Communication Studies and related fields and for professionals concerned with language and communication who are keen to extend their research experience.

More information about this series at
<http://www.palgrave.com/gp/series/14504>

Martha C. Pennington
Pamela Rogerson-Revell

English Pronunciation Teaching and Research

Contemporary Perspectives

palgrave
macmillan

Martha C. Pennington
SOAS and Birkbeck College
University of London
London, UK

Pamela Rogerson-Revell
English
University of Leicester
Leicester, UK

Research and Practice in Applied Linguistics
ISBN 978-1-4039-4235-7 ISBN 978-1-137-47677-7 (eBook)
<https://doi.org/10.1057/978-1-137-47677-7>

Library of Congress Control Number: 2018946548

© The Editor(s) (if applicable) and The Author(s) 2019

The author(s) has/have asserted their right(s) to be identified as the author(s) of this work in accordance with the Copyright, Designs and Patents Act 1988.

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Cover illustration: © ImageZoo / Alamy Stock Photo

This Palgrave Macmillan imprint is published by the registered company Springer Nature Limited
The registered company address is: The Campus, 4 Crinan Street, London, N1 9XW, United Kingdom

Preface

This book is the product of a collaboration between two pronunciation specialists, one educated and based in the United Kingdom (Rogerson-Revell) and one in the United States (Pennington). We got to know each other and our common interest and work in pronunciation as colleagues in the English Department of City University (then Polytechnic) of Hong Kong under the Headship of Professor Jack C. Richards in the 1990s, and since that time, we have remained in touch and kept interacting about our work. It was therefore natural that we became partners in this book project, first commissioned by Prof. Chris Candlin with Martha and later reconceptualized as a coauthored work combining our two different orientations and backgrounds and incorporating a wide range of knowledge and perspectives on pronunciation teaching and research.

We have written this book aiming to present a novel, state-of-the-art and issues-centered view of the teaching of English pronunciation that also connects teaching to research. There are many books available on pronunciation, including textbooks for teachers with practical teaching ideas and introductory books on phonology and phonetics. However, there is still little that has been written which brings together research and teaching or relates pronunciation to wider contexts. This book aims to fill this gap, helping teachers to see the relevance of research to teaching and presenting phonology in a wide-angle view as a crucial component of communication, identity, and the presentation of self.

We want to encourage and disseminate a view of pronunciation research and teaching, and of research and pronunciation practice more generally, as connecting in a two-way process in which research and practice function synergistically, in a trading relationship in which (i) the results of research inform practice and (ii) the contexts of practice provide sites for research and research results that inform future research. This creates an ongoing cycle in which practice, rather than evolving in relative independence from research, is continually referenced to it, thus creating applied knowledge. The synergy between research and practice also ensures that theory, rather than evolving in isolation from practice, evolves with it, at the intersection of research and practice, so that theory has applicability in real-life contexts.

The book takes a broad-based look at English pronunciation teaching and research in a twenty-first century context of widespread knowledge of English as a second or international language and changing views of the importance of pronunciation in language teaching and communication. It aims to situate pronunciation teaching and research within a wider context that includes language learning theory, language assessment, technological developments, and the broader relevance of pronunciation in both education and employment. The view of pronunciation that we present encompasses the production and perception of meaningful sound contrasts in English consonants and vowels as well as prosodic or suprasegmental contrasts in stress, intonation, and other features that contribute not only to denotative meaning, and so to intelligibility, but also to many aspects of pragmatic meaning (e.g., in expression of style, identity, stance, and politeness), and so to understanding in a larger sense and the impact that a speaker has on a listener. We also include discussion of voice quality and fluency as aspects of spoken language performance that are considered to be part of pronunciation, and consider the nature of accent and its place in pronunciation teaching. The contexts in which pronunciation is considered include language classrooms and many kinds of real-world contexts, from courtrooms, to doctor's offices and hospitals, to call centers. They also incorporate testing as an important aspect of pronunciation practice and research.

One goal of the book is to offer up-to-date information on these different aspects of pronunciation, as a form of continuing education and

inspiration for teachers and as directions for researchers. The book is aimed primarily at those who teach pronunciation or wish to teach pronunciation, both in-service and pre-service teachers, whether teaching in countries where English is the primary or a secondary language, and whether teaching classes specifically focused on pronunciation or not. It will also be of value to those in the research and testing communities with an interest in pronunciation, in addition to those who have a concern with pronunciation as a job-related issue for employers, employees, and customers. We believe that our book offers something for all of these audiences, and we hope all readers will find it informative, original, and interesting, in its broad scope, its up-to-date coverage, and the range of topics discussed.

The book is structured in eight chapters providing in-depth coverage with extensive and current references to literature. Chapter 1 addresses the nature of pronunciation in our broad conception and the types of meanings and functions it fulfills in communication, as we attempt to show that it has a greater importance in communication than is often realized, and so should command significant attention in teaching. Chapter 2 considers language learning with a focus on second language (L2) acquisition in instructed and uninstructed contexts and as contrasted with first language (L1) acquisition. Chapter 3 sets the teaching of pronunciation in a historical, theoretical, and international context and considers the factors that can be involved in making curriculum and teaching decisions about pronunciation and how these decisions might be influenced by research. Chapter 4 continues the focus on teaching by looking more closely at teachers and teaching approaches and methods. Chapter 5 is devoted to educational technologies and their potential for enhancing pronunciation teaching, learning, and assessment. Chapter 6 then turns to assessment and the many issues associated with the standardized testing of pronunciation as part of speaking proficiency or as a separate aspect of proficiency, with implications drawn for classroom-level assessment and for testing research. Chapter 7 considers the wider applications of pronunciation beyond the L2 speaking or pronunciation classroom, including in L1 literacy, speech therapy, and teacher education; in forensic linguistics, healthcare, and business and professional communication; and in the styling of speech in politics and social

communication. The final chapter, Chap. 8, offers a reconsideration of teaching and research in pronunciation and of the importance of continually relating research to practice and practice to research, and of cross-fertilizing different areas of knowledge.

London, UK
Leicester, UK

Martha C. Pennington
Pamela Rogerson-Revell

Acknowledgements

Martha would like to acknowledge the profound and enduring influence on her thinking and writing of the thinking and writing on phonology of Dwight Bolinger and David Brazil, each of whom she met only once but whose works she continues to benefit from studying, and of the thinking and writing in language learning and language teaching of Nick C. Ellis, Michael H. Long, and Jack C. Richards, all of whom she had the pleasure to work with for a period of time that set an ongoing interest in their ideas. She also owes continuing debts for both inspiration and knowledge of linguistics in general and phonology in particular to William Labov and Michael Halliday. Each of these scholars has influenced her contribution to this book through various stages of development.

Pamela has similarly been influenced by the many great phonologists and linguists who have helped take the field forward in recent decades and by the dedicated individuals, such as Richard Cauldwell, Judy Gilbert, John Levis, and many others, who have shown the importance of relating pronunciation research to pedagogic practice.

We also wish to thank two contemporary pronunciation specialists, José Antonio Mompéan González of the Universidad de Murcia and Jonás Fouz-González of the Universidad Católica San Antonio, for reading the preliminary version of our manuscript and giving detailed feedback on it. Their combined expertise in pronunciation theory, research, teaching, and technologies meant that they were able to provide us with

many kinds of perceptive input that we could use in revising our work towards the present version, while also drawing our attention to several relevant references that we have included in the final work. Further thanks go to the series editor, Jonathan Crichton, for his review and encouraging comments on our initial manuscript. If any errors of language or fact remain, these are of course our own.

In addition, we would both like to publicly acknowledge the value of putting together our differing but overlapping areas of experience and knowledge, and working out the common ground over the years spent writing this book. As admirers of each other's work with a long-term association, we had a good starting basis for conceptualizing and writing this book together—largely at a distance but with periodic face-to-face contacts—and for sustaining our joint efforts over a long period. The book created from our collaboration has evolved as a product of our mutual and highly interactive engagement, with equal effort on both sides. We believe that this collaboration has resulted in a much more original, comprehensive, and deeply considered book than either of us would have produced alone.

Contents

1	The Nature of Pronunciation	1
2	Phonology in Language Learning	57
3	Framing the Teaching of Pronunciation	119
4	Pronunciation in the Classroom: Teachers and Teaching Methods	173
5	Using Technology for Pronunciation Teaching, Learning, and Assessment	235
6	Assessing Pronunciation	287
7	Beyond the Language Classroom: Wider Applications of Pronunciation Research and Practice	343
8	Relating Pronunciation Research and Practice	401

Author Index 465

Subject Index 481

List of Figures

Fig. 1.1	Dimensions of pronunciation	7
Fig. 1.2	Not free at three	11
Fig. 3.1	Essential features of Jenkins' Lingua Franca Core (adapted from Jenkins, 2002, pp. 96–97)	137
Fig. 3.2	Nonessential features of the Lingua Franca Core (Jenkins, 2002, p. 98)	138
Fig. 3.3	Summary of “Amalgum English” phonological core (Cruttenden, 2008, p. 329)	139
Fig. 3.4	Summary of “International English” phonological core (Cruttenden, 2008, p. 333)	140
Fig. 5.1	Screenshot of acoustic feedback in <i>Pronunciation Power</i>	242
Fig. 5.2	Talking head <i>Baldi</i> (Ouni, Cohen, & Massaro, 2005)	253
Fig. 5.3	<i>Ville</i> screenshot—Pronunciation exercise	254
Fig. 5.4	<i>DEAL</i> screenshot	256
Fig. 5.5	Screenshot from <i>Tactical Iraqi</i>	258
Fig. 5.6	RALL with <i>IROBIQ</i> (left) and RALL with <i>ROBOSEM</i> (right) (Park, Han, Kang, & Shin, 2011)	260
Fig. 6.1	The CEFR phonological control scale (Council of Europe, 2001, p. 117)	315
Fig. 6.2	Extract from the CEFR phonological control descriptor scale (Council of Europe, 2017, p. 134)	317

List of Tables

Table 3.1	Factors involved in deciding pronunciation teaching priorities (adapted from Rogerson-Revell, 2011, p. 247)	154
Table 5.1	Some skill-focused non-CAPT resources	265
Table 7.1	Key to transcription conventions (Adapted from Rogerson-Revell, 1995, p. 244)	363
Table 8.1	Pedagogical research themes	403
Table 8.2	Summary of key findings of pronunciation effectiveness research	418



1

The Nature of Pronunciation

Introduction

Pronunciation is a much more important and pervasive feature of communication than is generally recognized. It is the crucial starting point for all spoken language, since thoughts must be articulated in sound in order to be heard and so to become a message that can be communicated to another person. Pronunciation is required not merely for talking, but for communicating and making sense to another person, that is, for making meaning in both an audible and an understandable form. A person's pronunciation ensures the clarity required for a listener to be able to pick out words from the stream of speech and put them together in meaningful, comprehensible patterns, and also projects information about the speaker and the context of communication that makes a certain impression and establishes the common ground between speaker and listener that is needed for effective communication. In both of these aspects, pronunciation is the foundation of **messaging** in speech—through articulating words and their combinations in grammatical and

discourse units and through projecting multiple facets of social and contextual meaning.

Research into pronunciation in real-world contexts, which today incorporate people's transglobal movements and interactions, is making its centrality and multiple functions in communication increasingly clear. A growing body of research demonstrates that pronunciation is an aspect of language and communication which demands attention in educational and workplace contexts where speakers who have different mother tongues seek to communicate in a common language, which in the world today is often English. The emphasis of this book is on pronunciation practice and research focused on teaching, learning, and using English in these real-world contexts of transglobal and international communication.

In this chapter, we take an in-depth look at the nature of pronunciation as a component of language and communication, in its many aspects as both production and perception of speech, and in its many functions for conveying meaning of different types. We begin by differentiating the terms and disciplines that are associated with the study of speech sounds, in order to make clear to readers our own references to pronunciation in this book. Next, we review the features of pronunciation and the different types of linguistic and social meaning expressed, first by the pronunciation of individual sounds and then by the pronunciation of stretches of speech. In that part of the discussion, we give many examples of the kinds of meaning conveyed by pronunciation and how misunderstanding may result from unclear pronunciation or different conventions for pronunciation and the interpretation of speech in different speech communities. That review is followed by a consideration of pronunciation as a feature of group and individual identity. The chapter also provides a review of key concepts as they are used in the different areas of pronunciation research and practice included in this book. By reviewing the multifaceted nature of pronunciation as a pervasive dimension of communication and introducing key terms and concepts for talking about pronunciation in its many manifestations, this introductory chapter lays the foundation for the remainder of the book.

The Nature of Pronunciation and Why It Is Important

A First Look at Phonology and Pronunciation

In linguistics, **phonology** refers to the **sound system** of a language, that is, the distinctions in sounds that are meaningful for that language, or to the sound stratum or level of language, as distinct from the other “higher” strata (e.g., of lexis and syntax) of language. Phonology can be thought of as the surface level, or the building blocks, of a language. All of the spoken units of a language, from syllables up to whole discourses, are expressed through or composed of speech sounds, **segmental features** or **phonemes** (consonants and vowels) and **suprasegmental features** or **prosodies** (properties of stretches of speech). Phonology is therefore one of the aspects that can be described or analyzed about a language and its individual elements (words, phrases, clauses, sentences, and discourses such as conversations or speeches). It is also one of the aspects of speech that can be described or analyzed with respect to individual speakers or groups of speakers.

Phonology comprises the meaningful units of sound out of which all spoken language is formed and connected, by convention, to meanings that human beings recognize and respond to—both internally, in terms of their thoughts and feelings, and externally, in terms of their interactive moves. Phonology can therefore be viewed as having both psychological and social dimensions. Phonology also has a cognitive dimension, since the articulatory, auditory, psychological, and social patterning of spoken language is imprinted in specific neural pathways. The brain is then able to control and integrate all aspects of phonological performance, both subconsciously and consciously, to ensure that speech is produced with a high degree of understandability according to the speaker’s intention.

Pronunciation is a prominent term among a number of different terms used within the realm of phonology and the various types of research and practice connected to the sound stratum of language. Although *phonology* is sometimes used as a cover term for all of the phenomena related to linguistic sound, it is often restricted to the description

or the study of meaningful distinctions in sound of a language, and on this basis differentiated from **phonetics**, which refers to the description or the study of the details of language sounds. Linguists regularly use these two terms with this contrast in mind, *phonology* to refer to the system and units of linguistic sound that are meaningful for a language and *phonetics* to refer to the physical properties of those units. The emphasis of theoretical linguists on **theoretical phonology** (or in some cases, **theoretical phonetics**) can be contrasted with the practical applications of applied linguists, which can be referred to as **applied phonology** (or in some cases, **applied phonetics**). The term *pronunciation* tends to have a practical or applied emphasis and so is generally not used by theoretical linguists and researchers in second language acquisition (SLA), who typically refer to *phonology* (or occasionally *phonetics*) as their area of study. Language teachers generally use the term *pronunciation*, referring to an area of proficiency in language learning or a type of skill in spoken language performance, rather than *phonology*.

Researchers and practitioners with a practical or applied emphasis may use any of these terms (*phonology*, *pronunciation*, or *phonetics*) together with others, such as **articulation**, relating to the mechanics of producing speech sounds (e.g., speech therapists), or **accent**, relating to the general characteristics of speech that are associated with a certain geographical locale or social group (e.g., managers and trainers in business). Social psychologists may refer to *pronunciation* or *accent* as a focus of investigation on people's attitudes to specific languages or speaker groups. Because we aim to focus on the practical aspects of phonology, we will refer to *pronunciation* for the most part, while using the other terms as appropriate for our coverage of research and practice in the various disciplines and areas of spoken language performance included in this book.

As a type of linguistic skill or language proficiency, pronunciation involves learning to articulate and discriminate the individual sound elements or phonemes making up the system of consonants and vowels of a language, sometimes referred to as **segmental phonology**, and the features of connected speech making up its **prosody** or **prosodic system**, sometimes referred to as **suprasegmental phonology**. The prosodic system or suprasegmental phonology includes, at a minimum, **tone** and **intonation** (defined by pitch), **rhythm** (defined by duration), and **stress**

or **accentuation** (defined by acoustic intensity, force of articulation, or perceptual prominence). From the perspective of language teaching, prosody may also include **articulatory** (or **vocal**) **setting**, a complex of specific postures of the vocal organs (lips, tongue, jaw, and vocal folds), and/or **voice quality**, the vocal characteristics resulting from such settings, that are associated with different languages and pragmatic meanings.

Phonemes are key to the makeup of words and their component parts—**syllables**, the allowable individual phonemes and phoneme combinations that can carry stress (e.g., /a/ alone but not /b/ alone; vowel [V] and consonant [C] in combination, /ba/ [C + V] and /ab/ [V + C]; and the vowel flanked by consonants /bab/ [C + V + C] and //blabz/ [CC + V + CC]). Individual phonemes differentiate rhyming pairs (e.g., *lap* and *cap*, *up* and *cup*, *seek* and *peak*) as well as all kinds of **minimal pairs**—pairs of words that differ in meaning based on a difference in one phoneme (e.g., *cab* and *cap*, *cup* and *cap*, *clap* and *cap*, *pick* and *peek*). Prosody comes into play when individual consonants and vowels are joined together to make syllables, as the components of the meaning-units (**morphemes**) composing words, which are the building blocks of phrases and all longer grammatical units and stretches of speech. Patterns of rhythm, stress/accents, tone, and intonation delimit the structure and meaning of words and larger units.

Intonation is sometimes referred to as **speech melody** or, informally, the “tunes” of language. Traditionally, American linguistics has made a distinction between *tone* as referring to word-level pitch patterns and *intonation* as referring to sentence-level or utterance-level pitch patterns (and often incorporating stress patterns as well) that is not made in British linguistics, where tone is a component of intonation (e.g., Halliday & Greaves, 2008). In this book, we will sometimes use *tone* to refer to pitch patterns or **contours** that function above the word level, reflecting the British tradition followed in some studies. As in the case of other terms connected to pronunciation teaching and research, we seek to avoid terminological confusion and overload while also aiming to accurately represent the way that terms are currently being used.

The sound system of each language is unique, built on specific distinctions in phonemes and prosodic features. Languages differ in the size of

their phoneme inventories as well as in the specific phonetic features that differentiate individual consonants, vowels, and prosodic patterns and their **cue value**, that is, the relative importance of specific phonetic features. While some languages have a small inventory of vowels (e.g., Hawaiian, Serbo-Croatian) or consonants (e.g., Cantonese, Japanese), others have a large inventory of vowels (e.g., Danish, English, Finnish) or consonants (e.g., Hindustani, Lithuanian). All languages have distinctive patterns of rhythm and intonation within their grammatical units, but languages differ in the prosodic basis of **lexical** (word-level) distinctions and patterning. While in some languages (so-called **tone languages**) tone (pitch levels or contours) is a defining feature of individual words and word combinations (e.g., Hausa, Thai), in others, stress is a defining feature at the word level (e.g., Arabic, English). The consonant and vowel phonemes and prosodic patterns of individual languages, the specific phonetic features of their phonemes and prosodies, and the cue value of the individual features will overlap but also differ to a greater or lesser degree. The areas of overlap in phoneme inventories and prosodic characteristics across languages provide a starting point for language learning yet at the same time can lead a learner to give insufficient attention to differences (see Chap. 2).

Figure 1.1 gives an overview of the many dimensions in which pronunciation functions in language and communication. As a multi-level and multi-dimensional phenomenon (see Fig. 1.1), pronunciation assumes great importance in communication: it is a major aspect of understanding and interpreting spoken language and speakers' intentions. Pronunciation is important not only for clarity of message and denotative meaning (the type of meaning conveyed in dictionary definitions of words), but also for subtleties of message meaning and connotation (the type of meaning conveyed by the associations of words in their contexts of use) and in conveying a certain impression of the speaker. Viewed as a communicational resource, pronunciation is a key aspect of communicative competence that goes far beyond being understood in the sense of speaking in such a way that the audience is able to recognize the words being spoken (i.e., intelligibility): it incorporates being understood in the broader sense of speaking in such a way that the audience is able to interpret many things about the speaker's nature and orientation.

Pronunciation is the initial layer of talk through which speakers construct and listeners decode and interpret linguistic signals, as an indicator of:

Focal Linguistic Units and Boundaries

- phonemes
- syllables
- words
- phrases
- clauses
- multi-clause units

Focal Information Units and Boundaries

- words and their component morphemes and syllables
- information units (e.g., phrases and clauses)
- key words in information units
- main parts of a discourse

Different Types of Information

- new vs. continuing topics
- background vs. foreground in a story line or topic
- turn continuation vs. turn transition points in conversation
- assertions (statements) vs. queries (questions) vs. demands (commands)

Pronunciation is also a major ingredient in first impressions and in the interpretation of people's meaning and intentions, as an indicator of:

Nature

- inherent characteristics (sex and age)

Nurture

- place of origin
- education level
- social or communal identity (e.g., ethnicity, social class)

Situational Positioning

- communicative role and position
- attitude towards the audience
- attitude towards the topic of speech

Fig. 1.1 Dimensions of pronunciation

Pronunciation is a cue to the speaker's origin, social background, personal and communal identity, attitudes, and motivations in speaking, as well as the role(s) and position(s) which the speaker is enacting in a specific communicative context.

Pronunciation is an important aspect of spoken language proficiency that includes speakers' strategic competence:

Strategic competence is the way speakers use communicative resources to achieve their communicative goals, within the constraints of their knowledge and of the situation in which communication takes place. [In all communication], pronunciation has pragmatic effects because of its function in the affective framing of utterances and in defining social and individual identity. Phonological competence has strategic value in terms of a speaker's ability to relate to and express affiliation with others in a particular social group or geographical area. It has value in terms of academic opportunity and other kinds of opportunities that might be open to a speaker who has a certain type of pronunciation or who has mastery of a range of varieties or styles. It also has value on the job and the job market in terms of being able to communicate competently with specific types of customers, in terms of the image the speaker conveys and the employer wants to promote, and in terms of the geographical range of customers that can be effectively served.... (Pennington, 2015, p. 164)

In these many different ways, pronunciation is a social and expressive resource that can be used in conjunction with other linguistic resources to convey many different kinds of meaning. The wider value of pronunciation and its application across many aspects of language and communication is a central concern of this book.

Phonology as Key to Understanding in Communication

People interpret speech within the whole context of utterance, which includes not only the physical and situational features of the setting in which an utterance occurs, but also the background knowledge and assumptions people bring to the setting of communication. The context

includes many types of background knowledge such as the participants' linguistic competence and cultural background, their knowledge and assumptions about people as individuals and as types, about the communication process in specific situations and, in general, about the world and how it functions. Differences in participants' observable characteristics, such as their mode of talk, can have either a positive or a negative impact on communication—as can any differences in purposes, preferences, and values that participants construe as relevant to the conduct and interpretation of talk.

Since pronunciation is a main factor in participants' identification of differences in background, perception of each other, and construal of the speech event, it has a major impact on interactive dynamics and the creation and interpretation of meaning. As a general rule, people process speech by first attending to global features that allow them to form initial impressions. These first impressions help to guide the process of interpretation by cueing the speaker's

- Affective state and attitude: compare *Thanks a lot* spoken with high pitch (suggests pleasure, sincere thanks) vs. low pitch (suggests displeasure, sarcasm);
- Background knowledge and assumptions: compare the tag in *My son Ben's a good boy, isn't he* spoken with rising tone (suggests asking to know) vs. falling tone (suggests seeking agreement).

In addition, global properties of speech in the way of prosodic information help listeners identify the structure of the utterance and locate linguistic units within that structure: compare *no one has* spoken on one intonation contour with linking across the three words (*no one/has*) vs. with two intonation contours and a break after the first word (*no/one has*).

A person's pronunciation in all its aspects—including the articulation of specific phonemes, words, and phrases as well as the prosodies of connected speech—is an important aspect of being understood as one intends. Pronunciation is first of all a crucial determinant of whether a person can be understood at all. Each language and **language variety** (or **dialect**) of a language has different pronunciation features which must be

mastered in order to be understood by those who speak that language or variety. A certain “threshold level” of pronunciation clarity or accuracy (Hinofotis & Bailey, 1980), according to the norms and experience of the audience determining what is understandable to them, is required for communication to take place. This threshold level of skill in pronunciation depends on achieving a basic knowledge of the sound pattern of the language and ability to **perceive** its phonemic and prosodic elements and distinctions, together with a certain level of skill and automaticity in the mechanics of articulation required to **produce** those elements and distinctions.

With the goal of maximizing meaningfulness and coherence, speakers generally supply multiple cues to meaning in the way of the particular words, expressions, and grammatical patterns they select and in the way of prosodic and segmental features of their speech. Such multiple cues offer a degree of redundancy that can aid a listener’s processing and understanding of spoken language. However, a language learner’s limited knowledge of the L2 reduces the options for supplying multiple and redundant cues to meaning, and a learner’s limited automaticity of production limits the ability to balance different aspects of utterance production simultaneously.

Segmental Level

Inaccurate pronunciation of individual vowels or consonants can sometimes be compensated by other message elements and cues in the surrounding context, but it can cause real problems in communication in some situations. For example, pronunciation confusions or lack of differentiation by international medical graduates (IMGs), such as between the words *breathing* and *bleeding* (Wilner, 2007, p. 14), are critical to patient health and might in some cases be matters of life and death (Labov & Hanau, 2011). Although not all miscommunication is so serious, as in the constructed example of Fig. 1.2, a lack of differentiation between one phoneme and another can easily interfere with understanding and can also lead to impression formation and triggering of stereotypes that may have other kinds of impacts on communication (as discussed further below).

[Mr. Karen, a Division Manager at an Australian subsidiary of an international company, has hired a non-native speaker of English as his secretary. She has just telephoned Mr. Stevens, an employee, to set up an appointment.]

Secretary: Hello, Mr. Stevens. Mr. Karen would like to see you tomorrow to discuss some matters relating to the budget planning meeting next week. Do you have some time in the afternoon?

Mr. Stevens: I'm not free.

Secretary: Not free. How about four o'clock, then?

Mr. Stevens: No good.

Secretary: How about before free?

Mr. Stevens: No, I'm not free before three or after three. I'm busy all afternoon.

Secretary: Oh, sorry! How about in the morning?

Mr. Stevens: Yes, any time in the morning is OK.

Secretary: How about 9:00 am?

Mr. Stevens: Fine.

Secretary: Your appointment with Mr. Karen is confirmed for 9:00 tomorrow.

Mr. Stevens: [He hangs up.]

Fig. 1.2 Not free at three

As this constructed example indicates, segmental mispronunciation or misperception may interfere with understanding and communicative purpose to a greater or lesser degree. In addition to potentially causing misunderstanding and miscommunication, segmental errors, substitutions, and nonstandard pronunciation can cause listeners to become distracted from the content of speech and focused on its form, in some cases, resulting in annoyance (e.g., Fayer & Krasinski, 1987) and/or “switching off” and avoiding further contact with a speaker (Singleton, 1995).

Beyond making it possible to understand what someone is saying, the way individual vowels and consonants are pronounced gives listeners useful information in the way of cues—often unintentionally but sometimes deliberately—to the speaker's background. Thus, a person who says the first vowel in *chocolate* and *coffee* in a certain way, as [uɔ], cues possible origin or residence in New York City or nearby areas of New York State and New Jersey. As another example, a person's pronunciation of the *t* in *water* as a glottal stop [ʔ] cues origin or residence in Britain, whereas pronunciation of the *t* in *water* as a flap [ɾ] cues origin or residence in the United States—though some young Americans are starting to have glottal stop in *water* and other words with *t* in **medial** (middle) position. People acquire different features of pronunciation depending on where they live and their age because of the specific groups of people they associate and identify with. People may also intentionally adopt features of pronunciation in order to express their social identification or affiliation with speaker groups.

Besides cueing where a person is from, the way the person pronounces individual sounds or words can also be indicative of other characteristics, such as level of education or social status. A well-known example of the connection to social status is one reported by Labov (1966), who researched the pronunciation of /r/ after a vowel (**postvocalic /r/**) in three New York City department stores: Saks 5th Avenue (a high prestige, high-price store), Macy's (a mid-level store in terms of prestige and price), and Klein's (low-prestige, low-price). He expected the sales clerks in those stores to differ in social status according to the type of store where they worked and also assumed that this difference would be reflected in their pronunciation of postvocalic /r/ as **rhotic** (with the /r/ articulated) or **non-rhotic** (with only a vowel articulation), which has been found to vary significantly by region and social class.

For example, postvocalic /r/ has a strongly rhotic pronunciation in much of the United States, though upper and middle class speakers in some coastal areas (e.g., Boston, Charleston, and Savannah), especially older speakers who have long roots in those areas, tend to pronounce words spelled with /r/ after a vowel in a non-rhotic way, lengthening the vowel (and sometimes altering its quality as well). The sentence, *Park your car in Harvard yard*, with all the *ar* words pronounced [a:] or [æ:], is

often given to illustrate this usage and its geographical and social associations. As another contrast, whereas accents in Scotland and Northern England are generally rhotic, accents in Southeast England are generally non-rhotic, and many Australian, Asian, African, and Caribbean varieties of English tend to be non-rhotic. In England, but not in Australia or other parts of the world, the non-rhotic pronunciation of postvocalic /r/ is historically associated with upper and middle class speech. In such cases, the non-rhotic pronunciation has a certain prestige. Where there are social class differences in use of one or another variant pronunciations of a phoneme, it is often found that people tend to employ the variant used by those of higher socioeconomic status in careful speech and that used by lower-middle class or working class speakers in less careful, spontaneous speech or casual speech.

Such differences in the regional and social significance of different pronunciations of postvocalic /r/ formed the backdrop of Labov's (1966) New York City department store study. Labov asked the store clerks where a certain item could be found that he knew was on the fourth floor, to try to get them to say *fourth floor*, in order to see if they pronounced the postvocalic /r/ in those words in a rhotic or non-rhotic way. Then he pretended not to have heard them and asked them to repeat what they had just said, as a way to elicit a more careful speech style. He found that the clerks were less likely to pronounce /r/ in the rhotic way the first time, when they were not paying attention to their speech, whereas they were more likely to give a rhotic pronunciation the second time, in careful speech. This was especially true for the final /r/ in *floor*. In addition, he found that rhotic /r/ was more likely the higher one went up the social scale, so that the Macy's clerks were more likely to have this pronunciation than the Klein's clerks, and the Saks clerks more likely than the Macy's clerks. Thus, Labov confirmed that in New York City, a person's pattern of behavior involving the pronunciation of /r/ was a linguistic cue or linguistic **marker** for the person's social class and also for whether the person was speaking in a **casual speech style** or a more **careful speech style** in which attention was focused specifically on clarity, that is, on pronunciation.

For speakers of a second language (L2), pronunciation gives an impression of their language competence, and may also give a generally positive

or negative impression of them in other ways. Sometimes, by mastering what is considered a difficult sound in another language—such as, for an English speaker, the German *ch* or the French or Spanish *r*—an L2 speaker can receive a positive impression from first language (L1) speakers of those languages. This may mean that L1 speakers of those languages might be prepared to spend more attention, time, and effort in communicating with that L2 speaker, thus aiding in the learner's process of acquiring the language and potentially making good social or professional connections as well.

Paying attention to details of pronunciation and learning to imitate L1 speakers well can pay off. One of the authors of this book (Martha) had this experience in learning Turkish, particularly in relation to words spelled with *e* (as in the word for “I” *ben*) and *r* (as in the word for “one” or “a” *bir*). She noticed that Turkish *ben*, although spelled the same as the English name *Ben* and pronounced that way by the other English speakers in her class, was pronounced by her L1 Turkish tutor, a graduate student from Ankara, with a vowel that was closer to the English word *ban*, involving a lower tongue position and more open jaw and mouth than for English *Ben*. She also noticed that the typical English pronunciation of Turkish *bir*, which was pretty much the same as the English word *beer*, had the vowel approximately right but not the final consonant, which was quite breathy and sounded like an *rr* trill (as in Spanish *perro* “dog” or *burro* “donkey”), but whispered. Once she noticed how Turkish *e* and final *r* differed from English *e* and *r*, she tried to imitate the Turkish pronunciations of *ben* and *bir*, both very common words, every time she spoke. She soon found her Turkish teacher and tutor, as well as Turkish students in her EFL classes, commenting on how good her Turkish was, even though she was only a beginner! This positive response motivated her to keep at her Turkish study.

L1 speakers often think that the L2 speaker who has mastered certain features of pronunciation is a better speaker of their language than may in fact be the case. Although this positive perception can cause problems when limitations in the speaker's L2 competence are revealed in communication, it is also an advantage in that L1 speakers are more likely to interact with an L2 speaker whom they think is a competent communicator. Thus, paying attention to pronunciation can have a significant communicative payoff that aids language learning.

Prosodic Level

Beyond the basic ability to perceive and produce phonemes and combinations of these to achieve a required threshold of intelligible speech, speakers are able to convey many other aspects of the meaning of a message in whole or in part through pronunciation, as summarized in Fig. 1.1. This includes prosodic features signifying the grouping, continuity, and focusing of information (e.g., which elements cohere or show discontinuities; what is the relative importance of elements) and the communicative function of a linguistic unit in terms of its grammatical status and pragmatic meaning (e.g., whether it is intended as a query, an assertion, or a demand; whether it is to be taken seriously or in jest). Prosody is also an important indicator of a speaker's attitude towards the audience, and may even determine whether a listener will give the attention and effort needed to receive and interpret the speaker's message. In these different ways, prosody contributes to a speaker's ability to convey and a listener's ability to comprehend meaning and intention. In Hallidayan terms, prosody, and specifically tone and intonation, can express *textual* (context-related) meaning, *ideational* (logical sequence) meaning, and *interpersonal* (social) meaning (Halliday & Greaves, 2008).

If the prosodic features of speech diverge from what a listener expects in a particular context, there can be misunderstanding, sometimes with serious consequences. For example, incorrect stress on numbers can cause misunderstanding between an air traffic controller and a pilot over whether the wind speed at ground level for take-off or landing is gusting to *fifteen* or *fifty* miles per hour: with stress on *fif-*, *fifteen* may easily be heard as *fifty*. Wilner (2007, p. 14) gives the example of a doctor's ability to clearly differentiate in pronunciation between *15 mg* versus *50 mg* as potentially critical to patient health. Somewhat less serious but nonetheless consequential in terms of misunderstanding and potential lost sales is the following example given by Tomalin (2010) of a transaction between a Filipino call center customer service representative (CSR) and a U.K. caller:

Customer: How much is the ticket?

Representative: FOURteen pounds.

Customer: FORTy pounds! That's too expensive. (p. 175)

Some of the problem in such cases may be the speaker's failure to articulate a final nasal in the *-teen* words, since native speakers will often shift stress in those words to initial position if the next word is stressed, as in 'fourteen 'pounds (Mompean, 2014), and yet may still be correctly understood to say *fourteen* and not *forty*.

Sometimes, with unfamiliar or unexpected prosody, there is no understanding at all. As the authors found when they were both living and working in Hong Kong, getting word tones wrong in speaking a tone language like Cantonese, in which minimal pairs often involve a difference in only a word's pitch contour, will usually result in complete communication breakdown. An example for English prosody is that of an EFL student studying in the United States who told the story in class of going to the supermarket and asking the cashier, "Where is the [lɛ'tu:s]?" (meaning to say *lettuce*). After being asked this question multiple times, the cashier became frustrated and refused to give the student any more of her time and attention, turning back to the other customers in line and telling the student he would just have to learn English so people would be able to understand him. When the prosodics are wrong, sometimes a listener is put off or just gives up. This is an example of the larger point that poor, incorrect, or nonstandard pronunciation can cause listeners to become annoyed and distracted from the speaker's message (Fayer & Krasinski, 1987), even to "switch off" and refuse to interact further with a speaker (Singleton, 1995).

On the other hand, an L2 speaker can often make up for limited knowledge of English by using prosody well. For example, it is possible for L2 speakers of French to significantly improve the response they will get from Parisians by adjusting their prosody on the universal greetings of (to a woman) *Bonjour, madame* or (to a man) *Bonjour, monsieur*. The prosody in question draws attention to the address term (*madame* or *monsieur*) through a large pitch contrast between the second syllable of *bonjour* and the address term (*madame* or *monsieur*), high pitch on the address term, and lengthening of the vowel in the final syllable. The highlighting of the word denoting the person addressed and the high pitch on that word and especially on the final syllable can be interpreted as a show of interest and politeness.

,bɔ̃\˘ʒuɐ̃ ,mɑˈdɑ:m

Bonjour, madame!

,bɔ̃\˘ʒuɐ̃ mɔ̃ˈsjø:

Bonjour, monsieur!

This is a type of prosody that can be considered empathic or exclamatory—the prosodic equivalent of *Bonjour, madame!* or *Bonjour, monsieur!*—and that also carries for Parisians a meaning beyond that of ‘Hello, madame/monsieur’ to include the sense of ‘Happy to see you!’ As a different kind of example, in Hong Kong, L2 speakers of Cantonese often find that when the tone pattern is right, L1 Cantonese speakers can understand even when the individual phonemes are not pronounced accurately. L1 Cantonese speakers also tend to respond more favorably to L2-accented Cantonese when the speaker has relatively good tones.

Miscommunication based on intonation can be serious in terms of the degree of misunderstanding and the inferences people might make from how something is said. Gumperz (1982) reported a clash at Heathrow International Airport in the 1970s between baggage handlers and recently hired Indian and Pakistani women cafeteria-line servers, who the baggage handlers said were treating them rudely. The newly hired cafeteria workers in turn felt that the baggage handlers were discriminating against them. Gumperz recorded and then analyzed interactions between cafeteria workers, both the newly hired Indian and Pakistani women and the older British women working on the cafeteria line, and their customers. He found a prosodic feature that differentiated the two groups of cafeteria workers that he claimed could be related to the bad feelings between the baggage handlers and the new cafeteria workers. He discovered that when customers came to the point in the cafeteria line where they had the option of gravy, the British servers would say the word *gravy* with a rising tone, in the conventionalized way of offering someone food, through a question signifying “Would you like some gravy?” In contrast, the Indian and Pakistani servers would say *gravy* with a falling tone, which came across to the baggage handlers as abrupt or surly, signifying not a politely voiced offer but more like an inappropriate command of “This is gravy, take it or leave it.”

A falling tone, which Brazil (1997) labeled a “proclaiming” intonation pattern, is conventionally employed in many varieties of English as a means of asserting something, whereas a rising tone, which Brazil (1997) labeled a “referring” intonation pattern, is a conventional means of suggesting or questioning rather than asserting. A person who uses falling intonation may be perceived not merely as making a proclamation or assertion, but also as assuming a position of controlling the discourse or the audience, whereas a person who uses rising intonation might be perceived as giving over control, or sharing control, of talk with the audience. These different positionings of the speaker by intonation will be perceived as appropriate and effective, or inappropriate and ineffective, depending on circumstances (Pennington, 2018b, 2018c).

As Cameron (2001) points out, when the Heathrow servers said *gravy* with a falling tone,

...it sounded like an assertion: ‘this is gravy’ or ‘I’m giving you gravy’—which seemed rude and unnecessary, since the customers could see for themselves what it was and decide for themselves if they wanted any.... But in Indian varieties of English, falling intonation has the same meaning as rising intonation in British varieties—in other words, there is a systematic difference in the conventions used by the two groups for indicating the status of an utterance as an offer. Since neither group was aware of that difference, the result is a case of misunderstanding. (p. 109)

Tannen (2014, p. 360) refers to this type of misunderstanding as a failure to understand the **metamessage**, “how you mean what you say” (p. 358) that is conveyed by intonation in its role of suggesting the context in which the message is to be understood. This is the important role played by intonation as what Gumperz (1982, p. 131) labelled a **contextualization cue**, a feature or set of features of **message form** intended by the speaker to guide a listener to a full understanding of **message function**, as a certain interpretation of the words used and their import in relation to context.

In the context in which Gumperz made his recordings, a server’s rising tone on the word *gravy* would likely be interpreted by a British English audience, or **addressee**, as a contextualization cue signifying a

metamessage of polite helpfulness and friendliness, indicating that the server was reaching out to the customer in offering gravy, and thus being customer-oriented, whereas a falling tone would not cue this kind of metamessage to British English speakers. Rather, it might be interpreted—especially in an intercultural encounter, where stereotyping can also come into play—as not showing an orientation to the customer, projecting unfriendliness and unhelpfulness, even hostility. Although it is likely that there are other contributing factors to the baggage handlers' perception of being treated rudely by the Asian cafeteria workers, not using the prosody which is customary and which the audience expects makes it harder to convey not only the intended meaning (the message), but also the politeness and helpfulness (the metamessage) that is conventional and so expected in dealing with customers in this and other similar contexts. By playing the recordings for the airport workers and pointing out the differences in tone and what each can signify, Gumperz helped the Asian and non-Asian employee groups see that they were working with different conventions regarding use of intonation as a contextualization cue and so to achieve some mutual understanding.

In this connection, Cruttenden (2014, p. 335) says that North Germans' tendency to use downward pitch glides (i.e., falling tones) can sound aggressive to English speakers, such as speakers of **General British English (GB)**, who use rising tones more and falling tones less. An essentially converse example is that in which statements ending in rising pitch (**high rising tone, HRT**) are interpreted to be questions, though they are not intended as such, or to be cues to the speaker's lack of conviction or insecurity in communication, though the speaker in fact neither lacks conviction nor is insecure in communicating. The phenomenon of using HRT in statements—so-called “Upspeak” (Bradford, 1997)—is a trend among young people in North America (both the United States and Canada), the United Kingdom, Australia and New Zealand, and India¹ that is intended to project a metamessage of friendliness and concern for the addressee's perspective, but is often misinterpreted or criticized by those (especially in the older generation) who do not use rising tone in this way. As a third example, research by Estebas-Vilaplana (2014) showed that mechanically manipulated pitch variation in the recorded Spanish and English versions of *wh*-question and answer sequences as produced

by a bilingual speaker elicited different responses from native Spanish and English speakers. Whereas the Spanish listeners judged the Spanish responses spoken with low pitch as polite, the English listeners judged most of the English responses spoken with low pitch as unexpected and rude, while judging the English responses spoken with a high pitch range as natural and polite. These are telling examples of how significant a person's pronunciation can be, intonation specifically, in the meaning conveyed to a specific audience.

Accent and Stereotyping

The general features of speech, including phonemes and prosody, give a certain impression of speakers and their status. Such general features are often labeled **accent** (see further discussion below). To give an example of how accent can convey different things, to some people, an American Southern accent signifies a charming or cultured person while to others it signifies low social status or lack of education (Campbell-Kibler, 2007). As another example, Americans typically think of people who speak with a standard British accent as charming, cultured, and educated, while Australians may consider those who speak with the same British accent as “stuck up.” On the other hand, not all British accents have these sorts of associations.

Linguistic stereotyping based on accent is a quick way to classify people. The same kinds of evaluations are applied as well to L2, “non-native,” accents. For example, people often say that English spoken with a French accent sounds emotional or romantic while English spoken with a German accent sounds unemotional or formal. Some linguistic stereotypes are quite negative and relate to marginalized social status, as is the case in Hong Kong for Filipino English (Lowe, 2000). These different responses to accented speech often stem from historical facts (e.g., that the British were the ruling class in America at one time or the fact that the majority of Filipinos in Hong Kong are in domestic service), or from characteristics of English as spoken with the features of a particular language (e.g., the prevalence of glottal stop and the lack of linking or **coarticulation** (also known as **sandhi**) between words in German-accented

English, which to an L1 English speaker may seem like emphatic or formal pronunciation). What is perceived as a **foreign accent** may be associated with negative and often unconscious stereotypes (Gluszek & Dovidio, 2010) as well as negative emotions towards speakers related to difficulties in understanding what they are saying. Problems in intelligibility may cause processing difficulty that makes listeners judge those they perceive as having a “heavy” foreign accent as less credible (Lev-Ari & Keysar, 2010).

Stereotypes that listeners connect to a person’s way of speaking can have significant and wide-ranging effects (see also Chap. 7). As Tannen (2014) points out, “negative stereotypes can have important social consequences, affecting decisions about educational advancement, job hiring, and even social policies on a national scale” (p. 372). In employment, a person may be discriminated against or considered to be disqualified for a certain job based on the person’s language variety or accent (Pennington, 2018b). Discrimination in employment, both hiring and advancement, is a well-known and widespread phenomenon in the case of African Americans who speak a distinctive, African-influenced variety that has been variously referred to as African American Vernacular English (AAVE), Black English, or Ebonics. John Baugh documents the negative “linguistic profiling,” which he defines as “the auditory equivalent of visual ‘racial profiling’” (Baugh, 2003, p. 155), that has dogged Black Americans based on their language and resulted in discriminatory practices in employment as well as in housing and other areas of life.

Another case of negative and discriminatory linguistic profiling can be cited in Hawaii, where there was a long-standing tradition that became increasingly prominent in the last quarter of the twentieth century of excluding teachers of Filipino background who were well-qualified in terms of their educational credentials and who were fluent English speakers from teaching in local schools based on “accent discrimination” (Chang, 1996, p. 139). School principals and members of the state Department of Education justified the status quo under the rationale that the majority of local students would not be able to understand or relate to the Filipino teachers. As another example of discrimination or profiling based on accent, in the early 2000s the state of Arizona justified assessing English teachers’ accents as a requirement for being a “qualified”

English teacher until this was challenged as violating teachers' civil rights (Ballard & Winke, 2017, p. 122). Even as people are exposed to more and more different varieties and accents of English through media, travel, and the global flows of migrants around the world, it seems that discrimination based on accent is alive and well, and may even be on the rise in both the United States and the United Kingdom, as Moyer (2013, p. 172) maintains.

Baugh (2003) also points out the converse, positive form of discrimination that is part of linguistic profiling, such as the favoring of white applicants for jobs or housing based on their "standard English" accent, or the favorable attitudes that Americans have of some L2 accents (e.g., British-accented English or French-accented English). Yet it must be pointed out that positive discrimination for some based on accent, such as a standard or prestige accent, automatically implies preferential treatment for them at the expense of discriminatory treatment of others.

Pronunciation as a Value-Added Factor in Communication

As this discussion has shown, pronunciation is not only a central and necessary aspect of communication to master, but in the best case is an aspect of spoken language that can result in positive interactions and add value and impact in aspects of life that depend on language and effective interaction with others. It is therefore an important basic as well as value-added factor for much of social, academic, and professional life centering on spoken language communication (as discussed further in Chap. 7). In the negative case, a person's pronunciation, of both individual phonemes and prosodic features, interferes with understanding what the person is trying to say (the message) and with interpreting what the person means (the metamessage). In the worst case, it can lead to serious miscommunication, misunderstanding, and negative attitudes and also be an aspect of negative and discriminatory linguistic profiling and the various types of social disadvantaging and discrimination that are associated with negative assessments of a person's language. Attitudes towards a person based on pronunciation are often the result of historical factors and stereotypes and so long standing and relatively automatic.

Pronunciation and Identity

People manage impressions by communicating their communal and individual identity in a variety of ways that they consider effective for presenting themselves and conveying their intentions to different audiences and for different purposes. “Language is central to speakers’ alignment with and against various role models and groups, as speakers project an identity by adopting linguistic features of those with whom they most associate and identify socially and psychologically” (Pennington et al., 2011, p. 178). The phonological conventions of different communities offer resources for individuals to project their affiliations as well as aspects of their identity (Zuengler, 1988) through the pronunciation of individual sounds, prosodic features, and accent. A certain type of prosody, such as the rising intonation in statements that is characteristic of “Upspeak” (Bradford, 1997), or pronunciation of a phoneme in a certain way, such as the pronunciation of Spanish *z* (e.g., *zorro* “fox”) and *c* (preceding *i* and *e*, e.g., *cielo* “sky” or *cebra* “zebra”) as interdental [θ] (e.g., by a Latin American Spanish speaker or North American English speaker), can be employed to intentionally project a certain image, affiliation, or identity to the audience.

A person’s pronunciation is an indicator of the identity and community membership(s) which that person claims and projects to others. Identity is something which is created dialogically, in interaction with others whom one associates and identifies with (Bakhtin, 1984/1929) in speech communities or “communities of practice” (Lave & Wenger, 1991; Wenger, 1998), which Wenger describes as “a group of people who share a concern or a passion for something they do, and learn how to do it better as they interact regularly.”² In a community of practice, specific knowledge and skills are valued and provide access and proof of membership. As Pennington (2018a) observes, “language learners can maintain a strong identity in one or more communities of practice where their primary language is dominant even as they also aspire to and cultivate status in one or more communities of practice in which a second language is dominant, such as a school, a language class, a Web community, or a multicultural group of friends” (p. 93). Language, and specifically pronunciation,

is a central aspect of identity that is tied to many other aspects of identity, such as country and region of origin, ethnicity, culture, education, and profession.

Since the language and specific variety or varieties of language which a person speaks communicate much about the person's identity and aspirations, and also provide social access and communicative power in specific communities and circumstances, learning a new language can affect the person's identity and opportunities:

Learning a second or additional language means acquiring a new way of communicating and presenting oneself that can open a person's identity to change, making identity more malleable and offering opportunities to experiment with new communicative features, such as accent or prosody, and with the social and cultural attributes of the new language and its associated discourses. It also means gaining access to new groups and communities of practice where new knowledge and behaviors can be developed that make it possible to participate in new discourses and to have a role in shaping those communities and discourses, thus enhancing a person's social and communicative power. Learning a new language can confer social status and can widen opportunities for education, employment, and new experiences that can impact identity. (Pennington, 2018a, p. 94)

These points define important positive aspects of language learning in general and pronunciation learning in particular that teachers need to be aware of and to consider with reference to the students they teach. At the same time, language teachers need to be aware that a learner's core identity, in being strongly interconnected to the learner's language and linguistic identity, may not be an easy thing to change and may even represent a felt threat to identity (Pennington, 2018a, p. 95).

As people seek to expand themselves and their experiences and opportunities by learning to speak an additional language, they naturally start from what they already know. Learning to speak a second language begins from a learner's identity, perceptions, values, and learned behaviors involving the mother tongue or L1, as connected to other aspects of the learner's identity, perceptions, values, and learned behaviors. The learner's L1 and the many associated areas of knowledge and identity provide a

cognitive, psychological, and social foundation for tackling the tasks of learning a new language as well as a perceptual basis for hearing another language and an articulatory basis for speaking it (see Chap. 2).

Key Concepts of Applied Phonology

Phonemes and Their Contextual (Phonetic) Variants

The sound system of a language consists of its individual *phonemes*, the distinctive consonant and vowel sounds of the language, and their **contextual variants** (or **allophones**), the specific pronunciations of the phonemes in different contexts. All of the speakers of one language or language variety share the same phonemes. Yet there is a tremendous amount of variation in the exact pronunciation of the phonemes of a language, in the way of positional variation of phonemes in sequence as well as regionally and socially conditioned variation. Phonemes not only have different pronunciations in different linguistic contexts, they also have different **regional variants** (variants associated with different regional accents) and **social variants** (variants associated with different social groups, such as male and female speakers, upper class and middle class speakers, and different ethnic groups, as well as with different speech styles, such as casual and careful speech). Often variant pronunciations signal sound changes in progress, with some variants representing older features of a language and others representing newer features which have been introduced into the community such as through popular media or new speaker groups (Labov, 2001) and which are spreading.

The different types of variant pronunciations of phonemes are **phonetic variants**; because they do not differentiate the meaning of words, they are not **phonemic**. Although the sounds of a language may be described or transcribed at a general (phonemic) level that does not include the detailed phonetic analysis of articulation in different contexts and for different speakers, if the focus is on social or regional characteristics, on the differences between languages or language varieties, or on

problems in communication stemming from pronunciation, it will often be necessary to pay attention to phonetic detail.

Listeners may respond differently to specific regional and social variants, as found, for example, in a series of investigations carried out by Campbell-Kibler (e.g., 2007, 2011) using a **matched guise** technique in which recordings of speech were digitally manipulated to provide pairs of speech samples that differed only in which of two phonetic variants occurred (e.g., an alveolar [n] or velar [ŋ] variant for *-ing*, or a fronted or backed variant of /s/). Respondents then rated the samples in terms of a series of adjectives for describing the speaker (e.g., relating to regional background or to characteristics such as intelligence or education). According to the results of the 2007 study, American listeners from both the South and the West were more likely to perceive an accent as “Southern” in the [n] guise for *-ing* and less likely to perceive it as “gay” or “urban.” Both studies showed that listeners rated speakers as less competent in the [n] variant guise for *-ing* and more competent in the [ŋ] variant. This result perhaps reflects the fact that the [ŋ] variant is more common in careful and middle-class American and British English speech while the [n] variant is more common in casual and working class speech (Labov, 1966, 2001; Trudgill, 1974). In addition, the 2011 study found that /s/-fronting caused listeners to rate male speakers as less masculine, more gay, and less competent.

As illustrated in these studies, variant pronunciations may be associated with a range of listener perceptions and evaluations of speaker characteristics, cueing geographical origin or urban/rural background as well as socioeconomic status, education, intelligence, competence, and personal characteristics. A variant that is associated with status or advantage in terms of education or economic power that would normally be considered prestige in a society may be labeled a **prestige variant**, such as the [ŋ] variant of *-ing* as contrasted with the [n] variant. The positive or negative status of regional and social variants is not such a clearcut matter, however, as a prestige variant may be negatively valued in some contexts, such as the [ŋ] variant of *-ing* if used in casual speech among friends (e.g., giving an impression of an inappropriately formal or careful style). Conversely, a variant that has some negative associations, such as [n] for *-ing* (Labov, 1966, 2001; Trudgill, 1974) or glottal stop for medial /t/

(Tollfree, 1999; Wells, 1982, pp. 324–325; Williams & Kerswill, 1999), can also have a kind of “covert prestige” (Labov, 1966, 2001) as signaling solidarity or membership in a specific group (e.g., a racial or ethnic group) or speech community.

Phonetics and Sociophonetics

Phonetics traditionally distinguishes the two branches of **articulatory phonetics**, which studies the physiology of speech and how speakers form, or articulate, individual sounds and combinations of these in longer utterances, and **acoustic phonetics**, which studies the properties of sound waves in speech and how these are perceived. Sometimes, a separate branch focused on speech perception is distinguished, that of **auditory phonetics**. The term **sociophonetics**, which came into widespread use starting in the 1970s in relation to Labovian **variationist sociolinguistics** (Foulkes, Scobbie, & Watt, 2010), is now used by many linguists to refer to the study of phonetic variation in speech that is socially meaningful, such as the differences in socioeconomic status and speech style conveyed by pronunciation of postvocalic /r/ (Labov, 1966), *-ing* (Labov, 1966, 2001; Trudgill, 1974), or medial /t/ (Tollfree, 1999; Wells, 1982, pp. 324–325; Williams & Kerswill, 1999) in various American and British speech communities; the social attributes (e.g., “urban,” “gay”) conveyed by the [ŋ] variant of *-ing* and by a fronted variant of /s/ (Campbell-Kibler (2007, 2011)); the differences in politeness and audience orientation conveyed by falling and rising pitch (Bradford, 1997; Cruttenden, 2014; Gumperz, 1982) and by high or low pitch (Estebas-Vilaplana, 2014); and, in general, the pronunciation features used in the construction or cueing (by speakers) or the perception and interpretation (by listeners) of individual or group identity (Hay & Drager, 2007). Intentional uses of prosodic and segmental phonology in communication as contextualization cues to metamessages are a type of behavior which, since it involves conventionalized social meaning associated with pronunciation, falls within the remit of sociophonetics.³

The full study of phonology requires some attention to phonetics, that is, to the details of sounds and how they are produced, including positional

variation and larger contextual effects on articulation, especially the different kinds of meaning conveyed by socially conditioned (i.e., sociophonetic) variation. In both the traditional sense of phonetics and the expanded sense of sociophonetics, the details of pronunciation and how those details are interpreted in terms of a speaker's meaning and intentions are especially important for L2 speakers as well as for language teachers and researchers.

Production and Perception

A person's pronunciation skill or competence has a mechanical aspect in terms of the functioning and control of vocal organs that is required for speech production. From this mechanical perspective, a person's knowledge of pronunciation involves the manipulation of the physiological organs forming a system of breath, resonance, resistance, and movement that makes speech possible. This is a system connecting the organs that control breathing (the diaphragm, the lungs, and the trachea) and that allow intake and outflow of breath (the mouth and the nose) with the articulators (lips, teeth, tongue, palate, jaw, pharynx, uvula, vocal folds) and their physical and mechanical properties. Speech is the result of a speaker's actions to manage and shape the air coming up from the lungs in complex ways that produce all of the variations in sound waves which people perceive as specific phonemes (consonants and vowels) and prosodic cues to meaning conveyed by pitch (tone or intonation), length or **duration** (rhythm), and volume or **amplitude** (loudness). Any or all of these features of prosody may contribute to meaning by cueing **prominence** in words (stress) and larger units (accentuated components of a message), the grouping of words into units, and the communicative function of utterances.

Speech is planned as a message that the speaker wants to get across and so is produced not with a goal of articulating individual sounds or words, but with a goal of putting those together to generate meaningful units and coherent stretches of spoken language. Because the generation of speech is usually performed in real time, with limited time to plan and produce an utterance that will convey the message which the speaker intends, there will be trade-offs between some aspects of message production as against

other aspects. If a speaker is able to put pronunciation on “auto-pilot,” that is, to set the articulators in a certain way and to speak according to well-established neuromuscular instructions and articulatory routines, the speaker will save cognitive and attentional resources for attending to other aspects of speech production. To the extent that they are able to do so, this is generally the path that speakers follow.

From the perspective of speech perception, a person’s pronunciation competence involves the ability to discriminate auditorially, through (the human faculty of) hearing, the consonant and vowel sounds (phonemes and phonetic variants) of a language and its other vocal signals of meaning in prosody. In order to be able to decode and understand what a speaker is saying, a listener needs to have skills of phonological perception that involve recognizing and decoding segmental phonemes and prosodic patterns and relating these to known words, grammatical constructions, and meanings, including a wide range of pragmatic meanings and communicative effects. Thus, in addition to being able to produce and perceive the segmental and prosodic components of the system, pronunciation competence in both production and perception requires a knowledge of the conventions linking features of pronunciation *form* to meaning and function, including how they can cue different contextual frames and metamessages (pragmatic and social meaning).

Pronunciation and Spelling

An **orthographic system**, which is a set of symbols for writing language down, incorporates a set of conventional symbols for spelling sounds and words. For many languages, the correspondences between phonemes and orthographic symbols (**graphemes**) are not one-to-one but many-to-one (i.e., different phonemes are spelled the same way) and one-to-many (i.e., one phoneme is spelled in different ways). The lack of correspondence between pronunciation and spelling is more extreme in some languages than others. For example, Spanish and Hawaiian have considerably less variation in sound-spelling correspondences than do French or English. The spelling of languages with a long history of writing and/or a long history of influence from other languages (e.g., English) is not a good

guide to pronunciation, nor is pronunciation a reliable guide to spelling. An example for English of one-to-many correspondence of a single phoneme to many different spellings is the homonym (**homophone**) set *I*, *aye*, and *eye* made from the vowel phoneme /aɪ/, which gives a taste of the various spellings of this vowel phoneme in English words, including not only those in these three words, but also, when co-occurring with at least one consonant: *y* and *uy* (*by/buy*); *ye* and *ie* (*dyeldie*); *ig*, *igh*, and *eigh* (*sign*, *high*, *height*); and *i* before a consonant with final silent *e* (*sine*, *hide*). An example for English of many-to-one correspondence in which different phonemes are spelled the same way is the various phonemes that can be spelled with a single letter *o*, such as in *do* /ʊ/, *done* /ʌ/, *bosom* /ʊ/ (first syllable) and /ə/ (second syllable), *co-operate* /o/ (first syllable) and /ə/ (second syllable), *people* (unpronounced or silent *o*).

Voice Quality and Articulatory Setting

As the sound system of a language or variety of language, phonology is tied to linguistic meaning and shared conventions which speakers draw on to communicate in an intentional way. Unintentional or unconventionalized vocal sounds, that is, those which do not signify consistent distinctions and patterns of meaning (e.g., involuntary cries in response to fear, pain, or shock; emotion-induced changes in pitch or other voice characteristics) are not considered to be part of language and so also not part of phonology. These are aspects of communication which, together with facial expressions and gestures, are often classified in the category of **paralanguage**. Learned and controllable vocal characteristics and segmental features can be employed intentionally as contextualization cues to project metamessages of different kinds through pronunciation, based on conventionalized prosodic and segmental patterns and their associated meanings, including situational affect and attitudes towards the audience (e.g., accommodating or condescending) and to other aspects of the speech event (e.g., sarcasm or irony), as well as aspects of personality and identity (e.g., friendliness or assertiveness, gender or class identity). Uncontrollable, natural physiological vocal responses which reveal a person's emotional and physical state (e.g., excited, relaxed) are

generally considered to belong not to phonology or pronunciation but to the domain of paralinguage. At the same time, natural emotive responses are not entirely separate from conventionalized prosody, as these are very likely the basis for conventionalized prosody such as the various meanings of raised pitch. Thus, the dividing line between phonology per se and what is considered paralinguage is not clearcut, and *paralinguage* is sometimes defined in a way that includes prosody, especially intonation.

Recognizable and identifying individual voice characteristics that in singing are referred to as **timbre**, in speaking are often referred to as *voice quality*, the characteristics of a given voice spanning stretches of speech (Laver, 1980). Van Leeuwen (1999) describes the dimensions of voice quality and timbre as tense/lax, rough/smooth, breathiness, soft/loud, high/low, vibrato/plain, and nasality. At an individual level, voice quality (e.g., a generally nasal and rough, or raspy, voice) and pervasive features of articulation (e.g., lipping) can differentiate and identify a specific speaker from others in the same speech community (though same-gender family members often have recognizably similar voice characteristics). Voice quality has also been associated by van Leeuwen (1999) and others (e.g., Yau, 2010) with different kinds of pragmatic meaning, functioning as a contextualization cue to metamesages. Yau (2010) gives the contrasting examples of customers' "hot anger," expressed by loud voice and/or high pitch (p. 117), versus "cold anger," expressed by soft voice and/or low pitch (p. 118). Pervasive articulatory features can likewise cue different kinds of pragmatic meaning and function as contextualization cues to metamesages, such as intentional lipping to signify childlike speech and hence naivete or silliness.

In addition, voice quality and "the physical postures of the articulators that produce a particular voice quality" (Pennington, 1996, p. 156) and a consistent shaping of articulation throughout speech, termed *articulatory setting* (Honikman, 1964) or *vocal setting*, have also been recognized as distinctive for different languages and varieties of language (Collins & Mees, 2013, p. 60). When articulators are set in a certain way, the articulatory setting provides a sort of mechanical or **motor template** for the production of speech that aids automaticity and fluent speech production. Specific articulatory settings—combining such features as the posture of the tongue (e.g., as fronted or backed, or as having the tongue tip

tapered or not), jaw opening (e.g., as relatively open or relatively closed), lip shape (e.g., as relatively spread, rounded, or neutral), and the posture of the vocal folds (e.g., as relatively tense or slack, or as shaping a certain type of glottal opening)—are associated with different accents and can identify a speaker's L1, language variety, or dialect. Collins and Mees (2013, p. 61) describe a range of articulatory settings involving tongue position as characteristic of different British English accents or varieties, and they contrast the articulatory setting of non-regional British English pronunciation (“loose lips, and relaxed tongue and facial muscles”) with that of French (“pouting lip-rounding, and tense tongue and facial muscles”). Although they are sometimes classified as outside phonology proper (i.e., as paralinguage), both conventionalized voice quality, as an indicator of pragmatic meaning, and articulatory or vocal setting, as the underlying articulatory basis of different languages, can be included within the domain of phonology or pronunciation.

Accent and Accentedness

It is a common misperception that speech can be accent-free, stemming from people's bias towards a familiar style of pronunciation. Munro, Derwing, and Morton (2006) define **accentedness** as “the degree to which the pronunciation of an utterance sounds different from the expected production pattern” (p. 112). As Pennington (2018b) notes:

People tend to perceive accent or accentedness in relation to a certain kind of pronunciation, which may be that of their own reference group or that of a standard language, as being the “normal”—that is, the common or expected—pronunciation or the “correct” pronunciation. When considered against the listener's pronunciation baseline or model, other pronunciations will be perceived as more or less divergent or marked, or more or less accented, in comparison to the unmarked pronunciation of the reference group or baseline, which appears to be unaccented or less accented.... Yet even the pronunciation of a standard language is accented: it is a standard accent rather than, say, a rural accent or a minority group accent.

Given the fact that what is the usual or expected pronunciation is entirely relative to the perceiver's concept of a standard or baseline, it can readily be seen that everyone has an accent.

A person's accent can be considered as those features of pronunciation that distinguish the person as coming from a specific country, region, or social group, including segmental as well as prosodic features. In the characterization of Moyer (2013):

Intonation, loudness, pitch, rhythm, length, juncture and stress are among accent's many features; all of which classify speaker intent as they encode semantic and discursive meaning: accent is a medium, through which we project individual style and signal our relationship to interlocutors. Even more broadly, it reflects social identity along various categorical lines. (p. 19)

Accent may not be the sum total of a person's pronunciation features but rather certain features of pronunciation which are more salient or distinctive as representative of a person's group origin or affiliation than others, and which may endure even as other features change. These may include regional or social variants, such as the rhotic and nonrhotic pronunciations of /r/ researched by Labov (1966) in New York City, or L2 variants, such as the different pronunciations of /r/ by German and Greek speakers of English researched by Beinhoff (2013). A speaker's origin may be detectable in features of accent even many years after changing geographical location or social affiliation. On the other hand, people often pick up new accent features from different places where they live.

As Levis (2016) notes, "[h]aving an accent that fits into a given social group may have benefits" (p. 154), such as the following:

- Cementing social bonds, as a key marker of social identity;
- Demonstrating social affiliation and so helping to gain access to social networks;
- Attributing the qualities of a leader to a person;
- Determining whether listeners will want to interact with a speaker and thereby affecting the availability of opportunities for language acquisition. (summarized from p. 154)

Yet, as Levis maintains,

...accent also has a dark side.... Four common consequences of accent (or even perceived accent) are isolation, an unfair burden on L2 speakers in communication, discrimination, and perceived social stigma. (ibid.)

Although accent can be distinguished from language competence, as a person with a detectable regional or L2 accent may be a highly competent speaker, linguistic stereotyping may nonetheless evaluate what is perceived as a “strong” accent as an indicator of limited competence in language and other things such as intelligence or education.

Accuracy and Fluency

Speech Production

Accuracy of pronunciation is a matter of articulating phonemes as intended so that they can be recognized by an audience as correct according to a certain system of distinctions between sounds. To develop high accuracy of pronunciation requires learning to both perceive and produce phonemes and their variants according to the norms of the community to which pronunciation is referenced. This may mean developing new targets for production that move away from inaccurate ones, such as those based on a different speech community or language, most especially a language learner’s L1 (see discussion in Chap. 2).

Being able to pronounce a language accurately is an automatic result of learning a language as L1 but not necessarily as L2. In the early stages of learning a language, L2 speakers will need to attain a pronunciation threshold, that is, a level of pronunciation accuracy which makes them understandable to others, while at the same time balancing all of the other demands of speaking. Once the threshold level is reached, learners are likely to focus on other aspects of speech production while backgrounding the achievement of full pronunciation accuracy as a goal. In the meantime, most L2 learners will draw on phonological similarities between the L2 and their L1 in applying their L1 categories, articulatory settings, and

pronunciation mechanics to production of L2 speech. The pronunciation of most L2 learners will therefore show, to a greater or lesser extent, inaccuracies that derive from applying the phonetic features and motor template of the L1 and from their lesser knowledge and automatization of L2 lexis and grammar, which creates a higher cognitive load in speaking.

Pronunciation accuracy is facilitated by focusing on the auditory features and the articulation of sounds both in isolation and in context, but a focus on pronunciation takes both time and attention away from other aspects of communication that more typically command speakers' attention. Speakers are therefore prone to automatize pronunciation to the greatest extent possible, following known and practiced routines for articulating phonemes and for realizing larger phonological patterns of coarticulation and prosody. These are generally based in a speaker's L1 and the specific varieties or dialects that the speaker commands. Engaging practiced articulatory routines and prosodic patterns, according to an automatized motor production template that allows a sort of "pronunciation auto-pilot" to function, can free up cognitive resources; and backgrounding pronunciation makes it possible to foreground meaning and lexical search in planning and producing speech in real time. At the same time, minimizing the conscious attention paid to pronunciation can reach a point of diminishing returns in terms of ensuring sufficient distinctiveness for understanding. For both L1 and L2 speakers, focusing on pronunciation accuracy may require explicit control that interferes with other aspects of message generation and slows down speech.

As Levinson (2000) observes:

[I]t is...possible to identify a significant bottleneck in the speed of human communication—a design flaw, as it were in an otherwise optimal system. The bottleneck is constituted by the remarkably slow transmission rate of human speech (conceived of as the rate at which phonetic representations can be encoded as discriminable acoustic signals), with a limit in the range of seven syllables or 18 segments per second.... (p. 28)

Levinson cites figures suggesting that the cognitive processes preceding articulation in speech production and comprehension take place three to four times faster than the rate at which a person is able to articulate speech. As he goes on to state:

It is this mismatch between articulation rates on the one hand, and the rates of mental preparation for speech production or the speed of speech comprehension on the other hand, which points to a single fundamental bottleneck in the efficiency of human communication, occasioned no doubt by absolute physiological constraints on the articulators. (ibid.)

Whereas drawing on L1 targets and mechanics for articulating the L2 results in some inaccuracy, it promotes fluent production while also making it possible to focus on other aspects of speech production involving the generation of meaningful lexicogrammatical units. Accepting a degree of inaccuracy in pronunciation may seem to be a reasonable trade-off for maintaining a focus on generating meaningful and coherent speech, though the degree of inaccuracy in individual phonemes, intonation, and other aspects of prosody can interfere with meaning and coherence.

Fluency, which is taken as an indicator of coherence and of highly proficient or “nativelike” speech in a second language, is often associated with notions of effortlessness, flow, or “fluidity” of speech (Browne & Fulcher, 2017, p. 38) and with notions of continuity and timing of speech (Dalton & Hardcastle, 1977; Lennon, 1990). As Browne and Fulcher (2017) have observed, “Fluency is as much about perception as it is about performance” (p. 37). Thus, discrete measures of **temporal fluency** having to do with the speaker’s timing of speech (Lennon, 1990)—including speaking rate, proportion of pausing, and presence of **disfluency markers** such as pause fillers or hesitators (e.g., *um*, *er*), false starts, and repairs—which are aspects of fluency that are quantifiable and measurable by humans or machines, may not equate to perceived fluency as a global or holistic measure.

Lennon (1990) noted that fluency could be defined broadly, as more or less equivalent to overall speaking proficiency, or narrowly, as a component of speaking proficiency that can be assessed separately. In Lennon’s (1990) view, fluency in the narrow sense can be taken to refer to the listener’s impression “that the psycholinguistic processes of speech planning and speech production are functioning easily and smoothly” (p. 391), exhibiting what Segalowitz (2010) labeled **cognitive fluency**, that is, “the efficiency of operation of the underlying processes responsible for the production of utterances” (p. 165). This narrow sense of fluency is what

Brumfit (1984) described as the “psychomotor” aspect of fluency encompassing speed and continuity. In this narrow way of conceptualizing it, fluent speech would be characterized by relatively long and continuous stretches of speech with relatively short stretches of silence (i.e., pauses) and relatively few pause fillers or other types of disfluencies such as false starts. Disfluent speech would be marked by short and discontinuous segments separated by long and/or frequent pauses, pause fillers, non-meaningful repetitions, stutters, incomplete thoughts, and other indications of nonautomaticity, difficulty, or loss of control while speaking.

This narrow conception accords with a common view of fluency as speaking without hesitation and of disfluency as hesitant or halting speech. According to Fayer and Krasinski (1995), the amount of pausing—measured by total pause time, percentage of pause time, and especially the length of the longest pause in discourse—result in listener judgements of speech as hesitant or not. Perceptions of a speaker’s hesitancy or overall continuity and discontinuity of utterance are a key factor for assessing spoken language performance as skilled or unskilled, native or non-native, normal or abnormal (as in disorders of fluency, see Dalton & Hardcastle, 1977, and Chap. 7). A speaker’s hesitations in speech are also a key factor determining whether a listener remains focused on the speaker’s message or becomes distracted and annoyed (Fayer & Krasinski, 1987). At the same time, in this narrow conception, fluency is defined without reference to meaning or the content of talk. For many purposes, this narrow definition would not suffice, since it does not differentiate fluent and meaningful speech from fluent but meaningless, incomprehensible, incoherent, or minimally informative talk.

Fluency is a concept associated with natural-sounding and non-hesitant speech produced in quantity and at a relatively quick rate. Fluency with a focus on pronunciation is a matter of coordinating segmental and suprasegmental aspects so that articulation occurs in a relatively continuous flow of speech, with few gaps or disruptions. Fluent production is focused not on individual sounds but on connected speech and so on units larger than phonemes, syllables, or individual words. At the level of phonology, fluency therefore is more about the prosodic than the segmental level although, contrary to the relatively strong stress features that come into play when a speaker is aiming for high accuracy,

fluent production often results in a rapid articulatory rate, with destressing and weakening of boundaries between syllables and words, and thus the extensive coarticulation that is normal in natural and automatized speech production. As Hieke (1985) observed, “Fluent speech is the cumulative result of dozens of different kinds of processes” (p. 140), including linking, levelling (i.e., assimilation), and “outright loss” (ibid.). It should be noted however that fluent speakers can vary speaking rate according to their intentions, such as slowing speech down to emphasize the part of a message expected to be new for the audience and to cue them to pay attention, or speeding through a stretch of speech to cue the part of a message that is expected to be received as “old news.”

To develop fluency at the level of pronunciation—that is, what has been called “**phonological fluency**” (Pennington, 1989, pp. 26–27, 1990, pp. 546–549) as distinct from fluency in a global sense that incorporates lexical and grammatical choices—a speaker must learn the ways in which sounds are preserved and altered in their connection with other sounds in context. Fluent speech processes are not necessarily the same from one language to another. For example, Delattre’s (1981) analysis of speech in four languages found a tendency in English for vowels to centralize towards the position of schwa [ə] under conditions of fluent production but no comparable tendency in the other languages investigated. Speakers who use articulatory sequencing or connecting strategies from the mother tongue may develop fluency in the second language with an L1 accent. Speakers who focus on the articulation of individual sounds or words in isolation, in contrast, may develop clear and accurate but non-fluent speech in a second language.

Speakers may be highly fluent and yet inaccurate in the sense that their pronunciation diverges in small or large degree from L1 or other speech community norms, making it hard for hearers from a certain speech community to understand speakers who originate in a different speech community. Pronunciation accuracy (or intelligibility, see below) according to “audience-determined norms is...an important goal, especially for those who must convey information to...native speakers, such as teaching assistants in undergraduate courses, supervisors in businesses or people who must speak to clients over the telephone in the target language” (Pennington, 1996, pp. 220–221). It is also important for L2 speakers communicating

with each other, since a certain degree of pronunciation accuracy is required for any communication to take place and, beyond that, to avoid miscommunication when the pronunciation of a particular phoneme or word is confused with another. As a different kind of problem, L2 speakers may have accurate articulation of individual sounds or words in the L2, yet lack the ability to put sounds or words together smoothly into longer continuous stretches of speech, thus making it difficult to communicate and also to capture and hold the attention of any audience.

Fluency, especially in the sense of temporal or phonological fluency resulting from automaticity (both cognitive and mechanical) in speech production, is a key goal in language learning as it makes it possible to focus away from articulation and more on meaning and other aspects of communication and the communicative context (e.g., speaker response) and also marks one as a competent speaker. Non-fluent, “over-hesitant speakers are likely to have difficulty communicating with native listeners for any length of time” (Pennington, 1996, p. 220), and other aspects of fluency are interconnected with phonological fluency, as discussed in Pennington (1990). Putting together words and syntax in coherent and meaningful grammatical units, displaying what can be considered *lexico-grammatical fluency*, presupposes a certain degree of phonological skill as a basis for fluent production. Conversely, the degree of mastery of syntax and lexis is a limiting factor on phonological fluency, that is, the ability to produce continuous speech in relatively long “lexical chunks” and grammatical units. Fluent speech is therefore in a basic sense discourse-level speech, and a language learner’s level of discourse competence, communicative competence, or overall proficiency is closely tied to the level of fluency in speaking an L1, in the case of a young child learner, or an L2, in the case of an older learner. Defined in the broadest terms, fluency in a language is equivalent to communicative competence or proficiency, and this broad competence or proficiency is strongly based in phonology.

Pronunciation accuracy and phonological fluency are important goals in speech that require a careful balance, and maximizing one in favor of the other can interfere with communication, especially (but not exclusively) in the L2 case. High pronunciation accuracy is often achieved by a speaker through a high degree of control of articulation of individual

sounds or words, resulting in relatively deliberate and effortful production that slows speech down and runs the risk of the speaker (and listener) losing the overall coherence of talk. A focus on accurate production of individual sounds or words is comparatively tiring, for both the articulators and the brain of the speaker, and may lose audience attention on message. There is a trade-off, however, in that rapid production of speech following a relatively automatized motor template, while it can produce high phonological fluency, can also lose articulatory distinctiveness, especially the specific quality of vowels and the details of articulation of consonants following vowels. As vowels and consonants become less distinctive and less distinct as words run together, individual phonemes and words may become less intelligible—a common outcome in both L1 and L2 interaction that leads to degraded communication, clarification requests, and attempts at repair that typically require moving away from rapid production of fluent speech.

It is therefore sometimes necessary for a speaker to deliberately abandon a speaking strategy in which pronunciation is on “auto-pilot,” following an automatized motor template or a goal of rapid or fluent production with a focus on meaning or semantics, and to switch to a more controlled pronunciation strategy, paying attention to language form and accuracy, in order to make sure that the audience understands what is being said. Speakers are especially likely to shift the focus in speech production away from meaning/semantics and towards language form and pronunciation accuracy when it seems that a specific addressee has not understood, as Labov (1966) illustrated in his New York City department store study demonstrating the tendency of store clerks to give a rhotic pronunciation of postvocalic /r/ when they thought their first production of *fourth floor* had not been understood. This is an important fact for instruction that aims to focus learners’ attention on pronunciation and its contribution to meaning (see Chaps. 3 and 4).

Speech Perception

Although accuracy and fluency are usually discussed as aspects of production, they can also be considered in relation to perception. Native speakers of a language who have normal hearing and intelligence can be

assumed to develop **perceptual accuracy**, the ability to recognize and differentiate distinct linguistic items, units, and patterns, and to link them to meaning, as well as **perceptual fluency**⁴ or **perceptual automaticity**, the ability to recognize and extract the form and meaning of linguistic items, units, and patterns quickly and with a minimum of processing effort. Given that phonology is the surface level of spoken language, perceptual accuracy starts with being able to recognize phonological segments and prosodic patterns, and perceptual fluency specifically incorporates skilled and rapid phonological processing. These kinds of phonological processing operate together with other kinds of processing skills drawing on lexical and grammatical knowledge as well as non-linguistic knowledge and processing skills (e.g., based on general knowledge and visual information in the context of speech) to generate an utterance interpretation.

For the L2 learner, perceptual accuracy and fluency will take time to develop and so utterance processing may be relatively slow and effortful, and may also be only partial, so that the learner needs to use inferencing skills and context to a high degree in trying to understand speech. In addition, learners' L2 perceptual processing routines will be based in part in the L1, so that L2 perceptual fluency—to the extent that a learner is able to achieve this—may be bought at a cost to L2 perceptual accuracy. Hence, the demands of rapid decoding and processing of L2 speech for meaning may lead to inaccuracies and errors (e.g., mishearing or misinterpretation of what is heard) or “dead ends” in the way of processing paths which do not result in a meaningful interpretation of a speaker's utterance.

Nativeness and Pronunciation Competence

As emphasized by a number of those working in applied phonology (e.g., Jenkins, 2000, 2002; Levis, 2005; Pennington, 2015), second language phonology and the teaching of pronunciation have traditionally been focused on **nativeness**, that is, a native speaker model for performance, as the goal of language learning and teaching. However, given the literal impossibility of being a native speaker of a language when one has in fact

grown up speaking a different language as mother tongue, and the difficulty (if not impossibility) beyond early childhood of developing L2 pronunciation that is indistinguishable from the pronunciation of a native speaker (see Chap. 2), the goal for L2 learners is usually stated as one of developing “nativelike” or “near-native” pronunciation. A great deal of the literature on pronunciation is couched in these terms.

Since the 1990s, applied linguists have been questioning the notion of the native speaker in relation to language teaching and L2 speakers in the context of **English as an international language** (EIL; e.g., Davies, 2003; Leung, Harris, & Rampton, 1997; Pennycook, 1995, 2012; Ricento, 2013; Ur, 2012). In Pennycook’s (2012) view, a more suitable criterion than nativelikeness would be that of a “resourceful speaker” (p. 99), meaning one who is “good at shifting between styles, discourses and genres” (ibid.). Pennington (2015) has described this kind of ability with reference to pronunciation in multilingualism or plurilingualism as competence in *multiphonology* or *pluriphonology*, involving speaker agency in using more than one language to express different aspects of identity and metessages, as in the practices of **style-shifting** (i.e., changing speech style or language to fit the context; Eckert, 2000; Eckert & Rickford, 2001), **crossing** (i.e., momentary use of a language from a group other than that to which the speaker belongs; Rampton, 1995), and **translanguaging** (i.e., use of two languages in combination; García, 2009). Ur (2012) observes that the majority of speakers around the world (outside America⁵) have competence in more than one language and that in the context of international English, the majority of those who speak English employ it as a common language, or **lingua franca**, to communicate with speakers whose mother tongue is not English. For this very large group of speakers, Ur maintains that a notion of language **competence** is a more appropriate concept for teaching than native speaker proficiency, and she suggests that such competence should be defined in relation to the communicative requirements of those who use English as an international language.

We suggest that a notion of **pronunciation competence** can perhaps be developed which considers aspects of communicative competence (Hymes, 1972; Savignon, 1983), both **receptive** (i.e., perceptual) **competence** and **productive competence**, that are specifically referenced to

segmental and prosodic phonology. Identifying those aspects of communicative competence that are specifically relevant to pronunciation can help to address the insufficient attention paid to pronunciation in both teaching and testing with an emphasis on communication (see discussion in Chaps. 4 and 6). As Ur (2012) emphasizes, such specifications may reference the competencies needed for English as used in international or lingua franca communication. As we would further emphasize, they might reference competencies needed for pronunciation performance in specific types of employment (see Chap. 7) and those related to identity, social meaning, and communicative pragmatics, including “pronunciation resourcefulness” and multilingual/plurilingual aspects of pronunciation in both perception and production.

Intelligibility, Comprehensibility, and Interpretability

Rather than defining it in terms of an external criterion or model of accuracy, nativeness or nativelikeness, or general pronunciation competence, an appropriate way of conceptualizing L2 pronunciation is in terms of **intelligibility**, which Munro et al. (2006) define as “the extent to which a speaker’s utterance is actually understood” (p. 112). A speaker may have an accent that diverges considerably from that of a native speaker yet be easily understood (depending on how strong the accent is perceived to be, as discussed above, and also how familiar the accent is, as discussed below). In the view of Isaacs and Trofimovich (2012), “in most situations of L2 use, what really counts is L2 speakers’ ability to be understood, rather than the quality or nativelikeness of their accent...” (p. 477).

As Derwing and Munro (2015) have observed, “In the last twenty years, ... both research and practice have placed a sustained emphasis on intelligibility, perhaps because there is now empirical evidence, first, that few adult learners ever achieve native-like pronunciation in the L2...and, second, that intelligibility and accentedness are partially independent...” (pp. 6–7). Jenkins (2000, 2002) has argued that mutual intelligibility among L2 speakers should be the main focus of their pronunciation, and she has proposed a set of minimum of phonological features required for mutual intelligibility between L2 English speakers—the **Lingua Franca**

Core (LFC)—with this goal in mind for pronunciation teaching (see Chap. 3 for details). Derwing and Munro (2005) agree with Jenkins “that mutual intelligibility is the paramount concern for second language learners” (p. 380), while also pointing out that

...ESL learners have to make themselves understood to a wide range of interlocutors within a context where their L2 is the primary language for communication and where, in many cases, [native speakers] are the majority. In addition, the purposes for communication may vary to a great extent when immigrants integrate socially in the target culture, which is an important difference from [English as an international language] environments. (p. 380)

Following Smith and Nelson (1985), intelligibility is one of three aspects or components of understanding in communication that can be recognized and differentiated: intelligibility, comprehensibility, and interpretability. Intelligibility, defined by Smith and Nelson (1985) in terms of word/utterance recognition (p. 334), “is interactional between speaker and hearer” (p. 333). Considered in information processing terms, *intelligibility* refers to the extent to which a listener is able to receive a message as it was intended to be sent and to decode its elements. Intelligibility can be considered a basic indicator of proficiency, in that a speaker must send a message in a certain form in order for a specific addressee or audience to be able to receive it clearly and without distortion. As mentioned above, from the point of view of pronunciation, intelligibility is linked to clarity and accuracy, which makes it possible for a listener to discriminate the message elements and recognize them as meaningful linguistic units—the words and the larger grammatical units composing the message. Intelligibility is also linked to fluency, as the processing of speech (by speaker and listener) can break down when continuity is disrupted. Thus, as Fayer and Krasinski (1995) report, hesitations correlate negatively with intelligibility.

In addition to a speaker’s ability to articulate as intended, intelligibility has to do with a listener’s ability to process the speaker’s utterance. Thus, like accuracy, intelligibility is in the eye of the beholder in the sense that it involves judgement of a speaker’s utterance by a listener—usually a specific

addressee or larger audience. Browne and Fulcher (2017), following Field (2005), observe that intelligibility has to do with how a listener processes the phonological content of a speaker's utterance, which relates to the listener's familiarity with the speaker's way of speaking, in particular, the speaker's accent. As Browne and Fulcher (2017) theorize, "increasing accent familiarity reduces the processing effort required for the phonological content of speech" (p. 40) and so makes that speech more intelligible to the listener. Accent familiarity can be considered to enhance perceptual accuracy and perceptual fluency, thereby speeding up the processing of speech. This is especially helpful when the speaker has heavily accented speech and/or is talking relatively fast, making it easier for the listener to keep up with the speaker's generation of message components in real time. How fast a person talks is an especially important intelligibility factor for a listener who has limited experience with the speaker's accent, or, in general, with the language the person is speaking, as the cognitive load of processing an L2 takes considerable effort and time. In addition, speakers articulate less distinctly when they are talking fast, and this can make segmentation of the stream of speech into individual words difficult for a listener, especially but not exclusively an L2 listener, thus impacting intelligibility and making comprehension difficult or impossible.

Comprehensibility, or "ease of understanding" (Munro & Derwing, 1995), is defined by Munro et al. (2006) as "the listener's estimation of difficulty in understanding an utterance" (p. 112). Whereas research has shown that the perception of accentedness is closely associated with segmental accuracy and other pronunciation factors (Saito, Trofimovich, & Isaacs, 2016, 2017), rather than with grammatical or lexical factors, comprehensibility is a multifaceted judgement that takes into consideration both segmental and prosodic features as well as temporal, lexical, and grammatical aspects of L2 speech (ibid.). It can be noted that hesitancy or disfluency is a factor in "ease of understanding" and hence in judgements of comprehensibility, as listeners can find that their comprehension suffers when a person's speech is substantially interrupted by pausing or other hesitators. On the other hand, "ease of understanding" can also be negatively affected by high fluency when it co-occurs with high accentedness, as listeners may experience moderate or even extreme difficulty understanding highly fluent speech if delivered in an unfamiliar accent.

Interpretability is the listener's ability to understand the speaker's intentions in terms of the communicative function or pragmatic force of the message, requiring functional and situational knowledge and knowledge of language-specific contextualization cues that signal metames- sages. Interpretability invokes not only speakers' and hearers' linguistic knowledge but also their social knowledge more generally. Although full interpretation of metames- sages and of the social function and pragmatic force of an utterance depends on the message being intelligible in terms of recognizing its components and comprehensible in terms of knowing the meaning of its words and grammatical structures, the interpretation of social or pragmatic meaning may be separate from, and may precede, analytical, item-by-item decoding and lexicogrammatical comprehension. Interpretation is a process which proceeds in a cyclical way, with global semantic and grammatical processing starting at an early stage and before decoding has been completed (Harley, 2008, p. 270). Utterance interpretation proceeds at the same time both bottom-up, from micro- level details in the speech signal, and top-down, from more global and higher level information—including the listener's knowledge of situa- tion, social meaning, and connotation—and cycles back and forth between processing levels. As in all of communication, the process of understanding is in part a guessing game—an inferencing or problem- solving process—of piecing together all of the clues or cues available in the utterance, the context, and the listener's stores of knowledge.

From the speaker's perspective, the process of building communica- tion starts at the opposite end of the sequence, that is, at the pragmatic and social level at which a message is contemplated and planned in rela- tion to context and audience, then built into semantic and grammatical units that are executed through a sequence of words, themselves sequences of phonemes or phonetic variants, framed by prosody (Levelt, 1989, 1999). Again, this makes speech production seem more orderly than it is, since speakers in most circumstances do not have time to plan fully at a global or macro level, much less at the micro level of fine details of lexical choice or articulation, before beginning an utterance. As a consequence, they start talking with a general intention of the message they want to communicate, but before they are sure what exactly they will say. This natural human tendency to jump into speech and build the elements of

the message while already engaged in talk is one source of errors, low intelligibility, and disfluency. Other sources of errors, low intelligibility, and disfluency are lack of time or attention and L2 status. Lack of comprehensibility can also be a sign that the speaker is talking without full knowledge (e.g., of topic, audience, context).

Concluding Remarks

As stated in the Preface, this book differs from other books on pronunciation in bringing together emphases on teaching and research, and in taking a much broader view of pronunciation than other works that incorporate a practical or applied orientation. The content coverage and orientation of this first chapter, in terms of the topics and concepts presented and our perspectives and emphases in discussing them, give an idea of where we are headed and the ways in which this book might differ from other discussions of pronunciation in language teaching, second language acquisition, and applied linguistics. As we have described and illustrated, pronunciation is a central aspect of human language that goes far beyond learning to articulate individual sounds, incorporating multiple layers of linguistic proficiency and types of communicative competence, including the ability of a speaker to produce and a listener to comprehend the meaning of message components relative to context, as well as expressing features of individual and group identity. Pronunciation can thus be considered from a wide variety of perspectives that can be explored, and have been explored, by teachers in their language classrooms and by researchers in educational and other contexts, and that offer a vast vista of further exploratory opportunities in teaching and research.

An important goal of this book is to present up-to-date information on these different aspects of pronunciation in a way that can provide inspiration, direction, and continuing education for teachers and researchers. With these goals in mind, it includes discussion of practical matters of curriculum and teaching, in addition to initiatives in the description and understanding of pronunciation that are offered from different fields of practical applications and research, and matters of theoretical interest

linking to foreign and second language learning and teaching, and to psychology and linguistics. Topics and concepts introduced in Chap. 1 will be the subject of further development in the chapters to come, based on the foundation laid in our definitions, illustrations, and discussion here. In Chap. 2, we build on the foundation established in this chapter to discuss the nature of language learning in both L1 and L2. These two chapters then form the basis for discussion of pronunciation research and practice in the classroom and larger contexts of society that are addressed in Chaps. 3, 4, 5, 6, 7 and 8.

Notes

1. For discussion, examples, and references, see Pennington, 1996, p. 155; Pennington, Lau, & Sachdev, 2011, pp. 184, 189–190.
2. <http://wenger-trayner.com/resources/what-is-a-community-of-practice/>
3. The concept of metamessage, which Tannen (2014) intends as a level or type of meaning different from that of strictly denotative or referential meaning, when connected to pronunciation suggests a distinction between phonology proper and *metaphonology*, or between phonology proper and some senses of *paraphonology* (see below), that may conflict with our broad conception of pronunciation in communication.
4. This is a term used in psychology to describe a subjective feeling of familiarity that results from “mere exposure” (often quite brief) to something, creating “perceptual fluency” in the sense of processing fluency or automaticity, and influencing people’s attitudes and affective judgements to what they have been exposed to (e.g., Bornstein & D’Agostino, 1994)—in particular, in a positive way (e.g., Reber, Winkielman, & Schwarz, 1998). Although this is rather different from the way we are using the term, we suspect that interesting connections could be made between *perceptual fluency* in our sense, which requires long and repeated exposure to a language, and positive attitudes to that language.
5. Ur (2012) relates the stereotype of America as the only country in which people can largely get by as monolinguals, though whether the extent of monolingualism is any higher in the United States than in Australia, New Zealand, or the United Kingdom can only be determined by empirical investigation.

References

- Bakhtin, M. M. (1984/1929). *Problems of Dostoevsky's poetics* (C. Emerson, Ed. and Trans.). Minneapolis: University of Minnesota Press.
- Ballard, L., & Winke, P. (2017). Students' attitudes towards English teachers' accents: The interplay of accent familiarity, comprehensibility, intelligibility, perceived native speaker status, and acceptability as a teacher. In T. Isaacs & P. Trofimovich (Eds.), *Second language pronunciation assessment: Interdisciplinary perspectives* (pp. 121–140). Bristol, UK and Blue Ridge Summit, PA: Multilingual Matters.
- Baugh, J. (2003). Linguistic profiling. In S. Makoni, G. Smitherman, A. F. Ball, & A. K. Spears (Eds.), *Black linguistics: Language, society and politics in Africa and the Americas* (pp. 155–168). Oxon, UK and New York: Cambridge University Press.
- Beinhoff, B. (2013). *Perceiving identity through accent: Attitudes towards non-native speakers and their accents in English*. Oxford and Bern: Peter Lang.
- Bornstein, R. F., & D'Agostino, P. R. (1994). The attribution and discounting of perceptual fluency: Preliminary tests of a perceptual fluency/attributional model of the mere exposure effect. *Social Cognition*, 12(2), 103–128. <https://doi.org/10.1521/soco.1994.12.2.103>
- Bradford, B. (1997). Upspeak in British English. *English Today*, 13(3), 29–36. <https://doi.org/10.1017/S0266078400009810>
- Brazil, D. (1997). *The communicative value of intonation in English*. Cambridge: Cambridge University Press.
- Browne, K., & Fulcher, G. (2017). Pronunciation and intelligibility in assessing spoken fluency. In T. Isaacs & P. Trofimovich (Eds.), *Second language pronunciation assessment: Interdisciplinary perspectives* (pp. 35–53). Bristol, UK and Blue Ridge Summit, PA: Multilingual Matters.
- Brumfit, C. (1984). *Communicative methodology in language teaching: The roles of accuracy and fluency*. Cambridge, UK: Cambridge University Press.
- Cameron, D. (2001). *Working with spoken discourse*. London, Thousand Oaks, and New Delhi: Sage.
- Campbell-Kibler, K. (2007). Accent, (ING), and the social logic of listener perceptions. *American Speech*, 82(1), 31–64. <https://doi.org/10.1215/00031283-2007-002>
- Campbell-Kibler, K. (2011). Intersecting variables and perceived sexual orientation in men. *American Speech*, 86(1), 52–68. <https://doi.org/10.1215/00031283-1277510>

- Chang, J. (1996). Lessons of tolerance: Americanism and the Filipino affirmative action movement in Hawai'i. *Social Processes in Hawaii*, 37, 112–146. Retrieved January 1, 2018, from http://efilarchives.org/pdf/social%20process%20vol%2037/sp37_chang_tolerance.pdf
- Collins, B., & Mees, I. M. (2013). *Practical phonetics and phonology: A resource book for students* (3rd ed.). Abingdon, UK and New York: Routledge.
- Cruttenden, A. (2014). *Gimson's pronunciation of English* (8th ed.). London: Routledge.
- Dalton, P., & Hardcastle, W. J. (1977). *Disorders of fluency and their effects on communication*. London: Edward Arnold.
- Davies, A. (2003). *The native speaker: Myth or reality*. Clevedon: Multilingual Matters.
- Delattre, P. (1981). An acoustic and articulatory study of vowel reduction in four languages. In B. Malmberg (Ed.), *Studies in comparative phonetics* (pp. 63–93). Heidelberg: Julius Groos Verlag.
- Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly*, 39(3), 379–397. <https://doi.org/10.2307/3588486>
- Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. Amsterdam and Philadelphia: John Benjamins.
- Eckert, P. (2000). *Linguistic variation as social practice*. Oxford: Blackwell.
- Eckert, P., & Rickford, J. R. (2001). *Style and sociolinguistic variation*. Cambridge: Cambridge University Press.
- Estebas-Vilaplana, E. (2014). The evaluation of intonation: Pitch range differences in English and in Spanish. In G. Thompson & L. Alba-Juez (Eds.), *Evaluation in context* (pp. 179–194). Amsterdam and Philadelphia: John Benjamins.
- Fayer, J. M., & Krasinski, E. (1987). Native and nonnative judgments of intelligibility and irritation. *Language Learning*, 37(3), 313–326. <https://doi.org/10.1111/j.1467-1770.1987.tb00573.x>
- Fayer, J. M., & Krasinski, E. (1995). Perception of hesitation in nonnative speech. *Bilingual Review*, 20(2), 114–121 <http://www.jstor.org/stable/25745268>
- Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly*, 39(3), 399–423. <https://doi.org/10.2307/3588487>
- Foulkes, P., Scobbie, J. M., & Watt, D. (2010). Sociophonetics. In W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.), *The handbook of phonetic sciences* (pp. 703–754). Oxford: Blackwell.

- García, O. (2009). *Bilingual education in the 21st century: A global perspective*. Oxford: Wiley Blackwell.
- Gluszek, A., & Dovidio, J. F. (2010). The way they speak: A social psychological perspective on the stigma of non-native accents in communication. *Personality and Social Psychology Review*, 14(2), 214–237. <https://doi.org/10.1177/1088868309359288>
- Gumperz, J. J. (1982). *Discourse strategies*. New York: Cambridge University Press.
- Halliday, M. A. K., & Greaves, W. S. (2008). *Intonation in the grammar of English*. London and Oakville, CA: Equinox.
- Harley, T. A. (2008). *The psychology of language: From data to theory*. New York: Psychology Press.
- Hay, J., & Drager, K. (2007). Sociophonetics. *Annual Review of Anthropology*, 36, 89–103. <https://doi.org/10.1146/annurev.anthro.34.081804.120633>
- Hieke, A. E. (1985). A componential approach to oral fluency evaluation. *Modern Language Journal*, 69(2), 135–142. <https://doi.org/10.1111/j.1540-4781.1985.tb01930.x>
- Hinofotis, F., & Bailey, K. (1980). American undergraduate reactions to the communication skills for foreign teaching assistants. In J. Fisher, M. Clarke, & J. Schacter (Eds.), *On TESOL '80: Building bridges* (pp. 120–133). Washington, DC: TESOL.
- Honikman, B. (1964). Articulatory settings. In D. Abercrombie, D. B. Fry, P. A. D. MacCarthy, N. C. Scott, & J. L. M. Trim (Eds.), *In honour of Daniel Jones: Papers contributed on the occasion of his eightieth birthday 12 September 1961* (pp. 73–84). London: Longmans, Green & Co., Ltd.
- Hymes, D. (1972). On communicative competence. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics* (pp. 269–293). Harmondsworth, UK: Penguin.
- Isaacs, T., & Trofimovich, P. (2012). Deconstructing comprehensibility. *Studies in Second Language Acquisition*, 34(3), 475–505. <https://doi.org/10.1017/S0272263112000150>
- 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. <https://doi.org/10.1093/applin/23.1.83>
- Labov, J., & Hanau, C. (2011). Pronunciation as life and death: Improving the communications skills of non-native English speaking pathologists. In B. J. Hoekje & S. M. Tipton (Eds.), *English language and the medical*

- profession: Instructing and assessing the communication skills of international physicians* (pp. 261–285). Bingley, UK; Leiden and Boston: Emerald; Brill.
- Labov, W. (1966). *The social stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.
- Labov, W. (2001). *Principles of linguistic change: Social factors*. Malden, MA: Blackwell.
- Lave, J., & Wenger, E. (1991). *Situated learning*. Cambridge: Cambridge University Press.
- Laver, J. (1980). *The phonetic description of voice quality*. Cambridge: Cambridge University Press.
- Lennon, P. (1990). Investigating fluency in EFL: A qualitative approach. *Language Learning*, 40(3), 387–417. <https://doi.org/10.1111/j.1467-1770.1990.tb00669.x>
- Leung, C., Harris, R., & Rampton, B. (1997). The idealised native speaker, reified ethnicities and classroom realities. *TESOL Quarterly*, 31(3), 543–560. <https://doi.org/10.2307/3587837>
- Lev-Ari, S., & Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology*, 46(6), 1093–1096. <https://doi.org/10.1016/j.jesp.2010.05.025>
- Levelt, W. J. M. (1989). *From intention to articulation*. Cambridge, MA: MIT Press.
- Levelt, W. J. M. (1999). Producing spoken language: A blueprint of the speaker. In C. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 83–122). Oxford, UK: Oxford University Press.
- Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. Cambridge, MA and London: MIT Press.
- Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, 39(3), 369–377. <https://doi.org/10.2307/3588485>
- Levis, J. M. (2016). Accent in second language pronunciation research and teaching. (Editorial). *Journal of Second Language Pronunciation*, 2(2), 153–159. <https://doi.org/10.1075/jslp.2.2.01lev>
- Lowe, C. T. (2000). *The outsider's voice: Discourse & identity among Filipina domestic workers in Hong Kong*. PhD thesis, City University of Hong Kong.
- Mompean, J. A. (2014). Stress shift in English: The case of teen numbers. In R. Monroy & I. Arboleda-Guirao (Eds.), *Readings in English phonetics and*

- phonology* (pp. 149–164). Valencia: Universidad de Valencia, IULMA (Institut Universitari de Llengües Modernes Aplicades).
- Moyer, A. (2013). *Foreign accent: The phenomenon of non-native speech*. Cambridge: Cambridge University Press.
- Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 45(1), 73–97. <https://doi.org/10.1111/j.1467-1770.1995.tb00963.x>
- Munro, M. J., Derwing, T. M., & Morton, S. L. (2006). The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition*, 28(1), 111–131. <https://doi.org/10.1017/S0272263106060049>
- Pennington, M. C. (1989). Teaching pronunciation from the top down. *RELC Journal*, 20(1), 20–38. <https://doi.org/10.1177/003368828902000103>
- Pennington, M. C. (1990). The context of L2 phonology. In H. Burmeister & P. Rounds (Eds.), *Variability in second language acquisition, Proceedings of the tenth meeting of the second language research forum*, Vol 2 (pp. 541–564). Eugene, OR: University of Oregon. Available at https://books.google.com/books/about/Variability_in_second_language_acquisiti.html?id=wkYqAQAIAAJ
- Pennington, M. C. (1996). *Phonology in English language teaching: An international approach*. London and New York: Longman.
- Pennington, M. C. (2015). Research, theory, and practice in second language phonology: A review and directions for the future. In J. A. Mompean & J. Fouz-González (Eds.), *Investigating English pronunciation: Trends and directions* (pp. 149–173). Basingstoke, UK and New York, NY: Palgrave Macmillan.
- Pennington, M. C. (2018a). Identity in language learning. In J. C. Richards & A. Burns (Eds.), *Cambridge guide on second language learning* (pp. 91–98). New York: Cambridge University Press.
- Pennington, M. C. (2018b). Pronunciation and international employability. Invitational Plenary, April 19, 2018. *AESLA (Spanish Association of Applied Linguistics) Conference XXVI*, Cadiz.
- Pennington, M. C. (2018c). Pronunciation pragmatics: The importance of intonation in communication. Invitational Lecture, January 30, 2018. School for Oriental and African Studies, University of London.
- Pennington, M. C., Lau, L., & Sachdev, I. (2011). Diversity in adoption of linguistic features of London English by Chinese and Bangladeshi adoles-

- cents. *Language Learning Journal*, 39(2), 177–199. <https://doi.org/10.1080/09571736.2011.573686>
- Pennycook, A. (1995). *The cultural politics of English as an international language*. London: Longman.
- Pennycook, A. (2012). *Language and mobility: Unexpected places*. Bristol, UK: Multilingual Matters.
- Rampton, B. (1995). *Crossing: Language and ethnicity among adolescents*. London: Longman.
- Reber, R., Winkielman, P., & Schwarz, N. (1998). Effects of perceptual fluency on affective judgements. Research report. *Psychological Science*, 9(1), 45–48. <https://doi.org/10.1111/1467-9280.00008>
- Ricento, T. (2013). Language policy, ideology and attitudes in English-dominant countries. In R. Bayley, R. Cameron, & C. Lucas (Eds.), *The Oxford handbook of sociolinguistics* (pp. 525–543). Oxford: Oxford University Press.
- Saito, K., Trofimovich, P., & Isaacs, T. (2016). Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels. *Applied Psycholinguistics*, 37(2), 217–240. <https://doi.org/10.1017/S0142716414000502>
- Saito, K., Trofimovich, P., & Isaacs, T. (2017). Using listener judgements to investigate linguistic influences on L2 comprehensibility and accentedness: A validation and generalization study. *Applied Linguistics*, 38(4), 439–462. <https://doi.org/10.1093/applin/amv047>
- Savignon, S. (1983). *Communicative competence: Theory and classroom practice*. Reading, MA: Addison-Wesley.
- Segalowitz, N. (2010). *Cognitive bases of second language fluency*. New York: Routledge.
- Singleton, D. (1995). Introduction: A critical look at the critical period hypothesis in second language acquisition research. In D. Singleton & Z. Lengyel (Eds.), *The age factor in second language acquisition* (pp. 1–29). Clevedon, UK: Multilingual Matters.
- Smith, L., & Nelson, C. (1985). International intelligibility of English: Directions and resources. *World Englishes*, 4(3), 333–342. <https://doi.org/10.1111/j.1467-971X.1985.tb00423.x>
- Tannen, D. (2014). Language and culture. In R. Fasold & J. Connor-Linton (Eds.), *Introduction to language and linguistics* (2nd ed., pp. 353–381). Cambridge, UK: Cambridge University Press.
- Tollfree, L. (1999). South East London English: Discrete versus continuous modelling of consonantal reduction. In P. Foulkes & G. Docherty (Eds.),

- Urban voices: Accent studies in the British Isles* (pp. 163–184). London: Arnolds.
- Tomalin, B. (2010). India rising: The need for two way training. In G. Forey & J. Lockwood (Eds.), *Globalization, communication and the workplace* (pp. 172–189). London: Continuum.
- Trudgill, P. (1974). *The social differentiation of English in Norwich*. Cambridge: Cambridge University Press.
- Ur, P. (2012). English as an international language: Implications for classroom teaching and teaching materials. Lecture at Bar-Ilan University. Retrieved January 1, 2018, from <https://www.youtube.com/watch?v=yTidAm0dRR0>
- van Leeuwen, T. (1999). Voice quality and timbre. In *Speech, music, sound* (pp. 125–156). London: Palgrave.
- Wells, J. C. (1982). *Accents of English 2: The British Isles*. Cambridge: Cambridge University Press.
- Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. New York: Cambridge University Press.
- Williams, A., & Kerswill, P. (1999). Dialect levelling: Continuity vs. change in Milton Keynes, reading and hull. In P. Foulkes & G. Docherty (Eds.), *Urban voices: Accent studies in the British Isles* (pp. 141–162). London: Arnold.
- Wilner, L. K. (2007). Communication skills training programs for doctors. *Academic Internal Medicine Insight*, 5(3), 14–15. Retrieved January 1, 2018, from AAIM Archives at <http://www.im.org/p/cm/ld/fid=209>
- Yau, J. N. W. (2010). Call centre discourse: Graduation in relation to voice quality and attitudinal profile. In G. Forey & J. Lockwood (Eds.), *Globalization, communication and the workplace* (pp. 107–124). London: Continuum.
- Zuengler, J. (1988). Identity markers and L2 pronunciation. *Studies in Second Language Acquisition*, 10(1), 10–49. <https://doi.org/10.1017/S027226310000694X>



2

Phonology in Language Learning

Introduction

Language learning can be seen as the flipside of language teaching, viewed from the perspective of the student or learner rather than the teacher, and as the backdrop or background for language teaching and language use that sets their potentials and constraints. In both of these senses, language learning stresses the contribution of the learner as a group representative and as an individual to the process of acquiring language, in particular, a second or foreign language, whether in a naturalistic or classroom setting. Research and theory on language learning provides a foundation of knowledge that can guide language teaching and assessment practices while also helping to elucidate other facts about language behavior. It also provides models for different kinds of research, quantitative and qualitative, that may have applicability in contexts of practice. Yet language learning research and theory are not always connected to language teaching and assessment, nor considered in the context of other areas of research and theory on language and communication, such as sociolinguistics or language in the workplace. This is just as true for pronunciation as in other areas of research and practice. This chapter therefore seeks

to provide a comprehensive overview of the topic of language learning, with an emphasis on phonology, that serves as a basis for making connections between this area of research and, on the other side, research and practice in language teaching and the wider context of pronunciation in other real-world contexts.

The chapter begins by taking a brief look at first and second language acquisition¹ to establish some basic facts and assumptions about language learning and to delimit some areas of difference and contrast in the L1 and L2 cases. We then review the nature of language acquisition in childhood with specific reference to phonology, followed by a review of second language acquisition with a focus on phonology. The latter section considers in detail the ways first and second language acquisition differ and the many factors which affect phonological learning in adolescence and adulthood. Throughout the chapter, we reference an extensive body of research and theory in first and second language acquisition, cognitive linguistics and cognitive psychology, and sociolinguistics to present a modern, empirically grounded view of pronunciation in language learning. The view presented incorporates considerations of production and perception and the multiple linkages of pronunciation to context that were introduced in Chap. 1, while highlighting individual differences that can account for differential success in pronunciation learning after childhood.

First vs. Second Language Acquisition

Language learning in childhood is a more effective process than language learning later in life, in the sense that children almost always become fully proficient in their mother tongue² whereas older learners almost never achieve full proficiency in a language. It used to be widely believed that the reason normal children learn their mother tongue perfectly whereas those who learn a language after early childhood learn it to varying degrees is that children learn language by means of an inborn language-specific cognitive capacity, a **Language Acquisition Device (LAD)** or **Universal Grammar (UG)** proposed by Chomsky (1965), while adolescents and adults learn language by means of their general

cognitive learning capabilities—their learning skills and general problem-solving abilities—which they can apply to any type of learning. Bley-Vroman (1988) referred to this difference between children and adolescent or adult learners as “**the fundamental difference hypothesis**” (p. 28, n. 1). Today, many linguists and psychologists discount the notion of an autonomous language acquisition faculty and maintain that both children learning their first language (L1) and adults learning a second language (L2) learn by means of general cognitive capabilities as applied to input gained from their environment and experience (see, e.g., Croft & Cruse, 2004; Ellis, 1996; Taatgen, 1999; Tomasello, 2003). An account of language learning based on cognitive processing of language input has become a mainstream view for cognitive linguists and cognitive psychologists as well as many (though not all) working in other branches of psychology and linguistics, including phonology (e.g., Ball, 2003; Browman & Goldstein, 1995; Bybee, 2001, 2010; Mompean, 2014).

Within a cognitive framework, a key reason for more successful language learning in early childhood is thought to be that L1 language learning is mainly **implicit** (incidental and automatic) learning of linguistic patterns bottom-up, from instances of meaningful input as it occurs naturally in communicative contexts, whereas L2 learning makes more use of top-down, **explicit** (intentional and controlled) modes of language learning outside of natural communicative contexts. In language learning, implicit learning occurs through processes of analyzing and storing information in the brain that are a basic mode of cognitive functioning, as listeners are able to keep track of acoustic details they encounter and of how frequently they occur, and then to utilize that information to discover patterns in linguistic input, such as words and phoneme categories (Bybee, 2001, 2010).

Without any intention to do so and without any awareness of it happening, listeners seem able to “maintain some sort of mental histogram” (Maye & Gerken, 2000, p. 530), like a large concordance registering instances of language they have heard, with the context of neighboring segments and their frequency of occurrence. This “mental concordance” is a statistical and probabilistic reckoning (Ellis, 1996) which is automatically stored in the brain as implicit knowledge in **long-term memory** and regularly updated based on new input. Learners are then able to use this

implicit knowledge in both speech perception, to comprehend and classify new input, and speech production, to realize their intentions and respond appropriately with language forms that suit the communicative demands of their current context. Explicit learning, in contrast, occurs through learners' intentional actions utilizing both long-term memory and the processing capabilities of **working memory** (Baddeley, 2012), including a processing component termed **phonological working memory** (or “the **phonological loop**”), together with the temporary information storage of **short-term memory**, to guide learning processes and outcomes, such as memorizing lists of words and grammatical paradigms or repeating recorded minimal pairs, phrases, or dialogues. Explicit knowledge and learned skills and strategies can aid in making less frequent, less obvious or **salient**, and more difficult or complex forms more learnable (Wulff & Ellis, 2018).

Many believe that the basic cognitive mechanisms which drive implicit learning are still operative in older learners, providing an important means of advancement in L2 learning. In the observation of Ellis (1996):

As learners' L2 vocabulary extends, as they practice hearing and producing L2 words, so they automatically and implicitly acquire knowledge of the statistical frequencies and sequential probabilities of the phonotactics of the L2. Their input and output modules for L2 processing begin to abstract knowledge of L2 regularities.... (p. 109)

However, adults may have “a somewhat weaker capacity for implicit learning, due particularly to age-related declines in the efficiency of *instance learning*” (Long, 2015, p. 41), that is, the ability to learn from individual instances of language input. These declines may stem from maturational changes, but they might alternatively show differences in cognitive processing due to experience. In the case of a child learning the L1, having a small number of stored instances of language use means that each experienced instance of language is likely to be processed as new information and so to have a strong influence on the developing language input and output modules, or templates, specifying instructions or routines for the perception and articulation of speech. In older learners, in contrast, input is less and less likely to be processed as new rather than of

a familiar type, that is, similar to what has already been experienced, processed, and stored in long-term memory. This may then account for a decrease in efficiency of instance learning.

There are certain things which young children can learn more easily than older children or adults, including spatial perception (vision), hand-eye coordination, normal movement (e.g., walking), and language. These are naturally learned in early childhood but difficult in adolescence or adulthood, such as for a person born blind who regains sight as an adult or cases like “Genie” (Curtiss, 1977) and the deaf child “E.M.” (Grimshaw, Adelstein, Bryden, & MacKinnon, 1998) who had very limited language learning as young children. These kinds of learning seem to involve what Lenneberg (1967) termed a “**critical period**” within which learning must take place to be successful (DeKeyser & Larson-Hall, 2005; Scovel, 2000). If there is such a critical period, or a number of critical or “sensitive” periods (Granena & Long, 2013b), these may reflect the influence of experience as much as or more than cognitive changes due to a genetically determined maturational schedule. Cases like “Genie” and “E.M.”—like those of older adult L2 learners who do not seem able to make much progress in language learning—may reflect the difficulty of changing long-established cognitive processing routines based on experience, which in the case of “Genie” and “E.M.” was many years of experience without much if any language input or output.

In the remainder of this chapter, we show how these facts about language learning apply with specific reference to pronunciation, looking at phonological learning first in L1 and then in L2.

Phonology in L1 Learning

Language Learning in Infancy: Getting to First Words

From birth, infants show abilities of perception and movement that are foundational for language learning, including “coordination between sensory modalities and between perception and motor action” (Thelen, 1991, p. 352), as well as imitative abilities which they use to imitate people around them (Thelen, 1991; Tomasello, 2003). According to Thelen (1991, p. 352),

newborns can coordinate auditory and visual information with their head and facial behaviors and will turn their heads towards sounds. In addition, “newborn infants as young as 42 min[utes] old will imitate adult mouth-opening and tongue-protruding gestures” (ibid.). This means that from the day they are born, infants are already directing their attention and natural abilities to learning language.

In the observation of Holzman (1997):

Babies utter non-cry vocalizations from the first day of life, but the young infant’s repertoire of sounds is quite limited. For the most part, the sounds are produced when the baby opens its mouth, and lets air out with no oral obstruction and with enough tension on the vocal cords to cause them to vibrate and produce sound.... Babies have to learn how to move their lips and tongues to make intentional sounds (p. 63).

Babies produce some sounds in their first six months that are not those of their native language. These “are mainly traceable to physiological factors, such as incomplete consonantal closure and natural physiological linkages of tongue and jaw position” (Velleman & Vihman, 2007, p. 26). At the same time, “children’s prelinguistic vocalizations as well as their speech perception show the effects of the input language” (Velleman & Vihman, 2007, p. 30).

Thelen (1991) described L1 phonological learning in terms of a series of stages which infants go through in learning speech that suggests the effects of both nature and nurture, in a developmental progression combining maturational changes with learning based on the ambient language, beginning with a “phonation stage” (p. 346) of making speechlike comfort sounds and then moving by the third month into a “gooing stage” (ibid.) of making guttural sounds and combining these with vowel-like sounds. The period of four to six months Thelen (1991) describes as an “expansion stage” (p. 346), in which “infants produce a variety of new sounds like raspberries, squeals, growls, yells, whispers, isolated vowel-like sounds, and immature syllables” (ibid.), before producing “canonical babbling” in the period of 7–10 months, when children babble strings of reduplicated open (CV) syllables made up of certain well-formed consonants (e.g., [d] and [m]) and “fully resonant vowels” (p. 346) that show

the influence of the ambient language. The canonical syllables observed in children's babbling, which Oller (2000) characterizes as "the prime rhythmic units in natural languages" (p. 6), represent an important advance in the infant's prelanguage repertoire. According to Oller (2000), canonical babbling by means of "the well-formed syllables of the later stages of infancy do not appear to be accidental, but instead show signs of intentionality; babies produce them repetitively, imitatively, and in playful social interaction" (p. 5, n. 2). Infants at this stage can therefore be seen as developing their interactive and communicative skills in relation to the sounds and rhythm of natural language in general and the sounds and rhythm of their L1 in particular. The stage of canonical babbling often leads to a "variegated babbling stage" of more complex babbling of strings of different syllable types (Thelen, 1991, p. 246) before children speak their first words, typically near the end of their first year.

Before they speak their first words, infants' speech perception has been neurologically set, or **entrenched**, channeling attention away from certain phonetic cues and properties while focusing attention on others, as a form of **perceptual narrowing**. As Velleman and Vihman (2007) observe, "Over the first year, infants develop familiarity with the commonly occurring prosody, consonants, vowels and consonant-vowel sequences of the ambient language" (p. 37). Specific words, often frequent ones, offer exemplars of individual phonemes and contextual variants which then become the reference point or model, often referred to as a **prototype**, for the learner's initial knowledge of where that phoneme is located in acoustic space. Specific phrases that are prominent in input to the child likewise serve as prototypes for prosody. Iverson and Kuhl (1995) and Kuhl and Iverson (1995) speak of these prototypes as exerting a "**perceptual magnet effect**" that influences learners' perception of speech sounds to be heard in terms of already established prototypes. The entrenchment of the prototypes or central exemplars means that perceptual sensitivity to sounds that are acoustically similar to the prototypes is reduced.

According to the **Native Language Magnet** (NLM) model, children progress in a "developmental sequence from universal perception of sounds," based on our mammalian genetic endowment, "to language-specific perception" in the period between 6 and 12 months of age, as "infants' mapping of ambient language warps the acoustic dimensions

underlying speech, producing a complex network, or filter, through which language is perceived” (Kuhl, 2000, p. 11854). The strong influence of the prototypes also affects the child’s production of speech, as models for how speech is supposed to sound. The organization of perception into such prototypes or central exemplars simplifies the learning process in providing a perceptual model and target that focuses attention for perception and simplifies the motor plan for production. At the same time, it means that the immature learner can hold something constant and keep cognitive load at a manageable level. The child then gradually accumulates the variation that is possible around each central exemplar, fine-tuning linguistic categories to take account of special contextual effects and variation in the speech community. The same principle is used for acquisition of L1 sounds, words, and grammar.

Once the perceptual warping of the child’s acoustic space has occurred, it is very difficult to modify, even if an L2 is learned in childhood:

The language-specific filter alters the dimensions of speech we attend to, stretching and shrinking acoustic space to highlight the differences between language categories. Once formed, language-specific filters make learning a second language much more difficult because the mapping appropriate for one’s primary language is completely different from that required by other languages. Studies of adult bilinguals, who were exposed to their second language after the age of 6, demonstrate magnet effects only for the first language, illustrating the potent effects of early linguistic experience. (Kuhl, 2000, p. 11854)

In Kuhl’s (2000) view, the L1 pronunciation filter or template may be established as early as age 1, that is, by the time children start speaking their first words.

The acquisition of the L1 phonological system thus involves amassing a great amount of detail about individual events of language use and then gradually reinforcing and strengthening some features—those which are most frequent and most salient or contrastive in the input and so most useful for their cue value—while weakening or **pruning** (Jensen, 2015, p. 60) the less frequent and less salient/contrastive features as low in cue value. The child’s L1 phonological system is therefore constructed empirically from a very large amount of ambient language data in the form of

input apprehended in fine-grained detail, analyzed by induction, and fine-tuned according to a criterion of cue value, that is, **communicative utility** or **functionality**. In L2 learning, it is rare for these characteristics of L1 learning to apply, and in virtually all cases, one or more of the conditions of successful phonological learning are missing.

Children's production of speech is built on and continually referenced to their perceptual template, which, having been formed by hearing and analyzing input from accomplished ("expert") speakers of the L1, represents a model for speech that is beyond what early language learners can themselves produce. The tuning by the child of vocal production to speech perception is a gradual process that shows the influence of both universal articulatory constraints in favoring certain sounds—the **unmarked** types, those which are easiest to produce and are most common across languages—over others—the **marked** types, those which are hardest to produce and are least common across languages—and the specific language to which the child has been exposed (Velleman & Vihman, 2007, pp. 26–27). Thus, for example, infants favor unmarked sounds (e.g., in the case of consonants, dental or alveolar stops, nasals, glides) and open CV syllable types in their babbling and first words, and they produce articulatory simplifications of marked L1 sounds they have heard in the input language as unmarked sounds in their own output (e.g., [w] for /r/, [d] for /ð/). Through much practice in babbling and later pronouncing words and longer utterances, children's prespeech vocal productions are gradually tuned to their more developed perceptual capabilities, which have already been tuned to L1 speech.

Experiments have shown strong, statistically based auditory pattern-finding skills in 7- and 8-month-old infants (Marcus, Vijayan, Bandi Rao, & Vishton, 1999; Saffran, Aslin, & Newport, 1996). Children retain in memory "very specific auditory traces" that include "sociophonetic aspects such as voice quality" (Velleman & Vihman, 2007, p. 36) and other contextual details relating to function, which form a mental database for analysis and learning. The specific language that the child is exposed to determines the child's linguistic knowledge and future linguistic performance. What is frequent in the input to the child becomes a central part of the child's linguistic system and the prototype or standard for the child's analysis and synthesis of speech. What is infrequent

becomes part of the child's peripheral or secondary linguistic system, to be considered when the more frequent or prototypical feature does not seem to apply in a specific case. What never occurs cannot in any sense affect the child's developing linguistic system and is ruled out by default, so that "what is not heard is taken to be impermissible in the language" (Velleman & Vihman, 2007, p. 37). It is thus lack of experience that rules out certain forms and structures as anomalous or uninterpretable, rather than explicit negative evidence such as corrective feedback.

In the first days of life, infants are not only tuned to their mother's voice, but also to the prosody of the conversational speech to which they are exposed and to the exaggerated prosody of caretaker talk, or **infant-directed speech** (Velleman & Vihman, 2007, p. 27). Infant-directed speech emphasizes language form and message content through exaggerated prosody and articulation in order to catch a child's attention and help the child to notice the important features of language and the surrounding context. In this way, the adult makes language more accessible to the child, more perceptually salient, and more memorable so that the child can focus on the important aspects of language and in this way understand and learn them. As Holzman (1997) observes, "Exaggerated intonation and high pitch are effective in directing and holding infants' attention" (p. 63) and are therefore common in caregivers' speech to children across languages (Fernald et al., 1989). Repetition and interaction aid in building up the child's perceptual database of instances of language and learning the different phonological forms that words may take. Moreover, as Kuhl (2000) states, "Mothers addressing infants [using exaggerated speech] also increase the variety of exemplars they use, behaving in a way that makes mothers resemble many different talkers, a feature shown to assist category learning in second-language learners" (p. 11855). Infant-directed speech has the additional purpose of tailoring the content of speech to the child, simplifying and focusing message content and lexis and using repetition to make it easier for the child to understand. It is also designed with a focus on interaction between the child and the caretaker, requiring a give-and-take that often incorporates play or game-like features and encourages learning by imitation. The special register of infant-directed speech appears to be designed specifically to foster interaction and language learning in context. In the view of Kuhl

(2000), for the child language learner, “vocal imitation links speech perception and production early” in life (p. 11856).

The ultimate result of all this input, interaction, and practice, occurring concurrently with many kinds of enabling maturational changes, is the child’s production of what others recognize as language, together with the development over time of an entrenched L1 phonological system connecting the brain and the auditory and vocal systems to the lexico-grammatical systems of language in complex ways that make possible the perception and production of speech. This L1 phonological system incorporates various types of complex neurolinguistic programming in two interrelated networks, forming:

- a *perceptual template* for discriminating sounds and recognizing words and larger meaningful patterns in the stream of speech; and
- a *motor template* for planning and executing articulation in coordination with the generation of meaningful lexicogrammatical units.

Identity, Variation, and Individual Differences in L1 Learning

A child develops an identity as an individual and as a member of a number of social groups, as indexed by language in general and pronunciation in particular (Labov, 2001). For preschool children, the family group is usually the primary group of reference whereas for school children, the peer group becomes an increasingly strong focus of affiliation. As they mature, children seek to develop an individual identity in terms of attributes which they value, learned as part of social behavior. A healthy sense of self develops in relation to an image of others, as the child seeks to be like or unlike those around him. Belonging to a social unit or group such as the family unit or a group of neighborhood and school friends is important at every age, as is the development of individual distinctiveness. Distinctiveness and individuality become increasingly important as a child matures into an adolescent. From adolescence to adulthood, the individual progresses to become an increasingly independent human being.

Kerswill and Shockey (2007) maintain that preschool children experiment with phonetic variation (p. 62), including different accents and degrees of formality, and can also vary characteristics of their speech in response to listener needs (p. 66). The kinds of variation and differences in speech styles which children are exposed to will not be identical. The influence of the language and the specific people around them, coupled with individual differences in response and in cognitive, learning, and speaking styles, continue throughout childhood, including adolescence, when speakers acquire features of speech as an aspect of establishing their peer identifications and affiliations, and of developing their communal and individual identity. It is therefore not accurate to say that language learning, whether viewed from the point of view of input or output, is uniform for all children. Instead, every speaker develops a linguistic system that overlaps to a substantial extent, while also differing to a greater or lesser extent from, that of others who speak the same L1, depending on: (i) the specific database of instances of language they have been exposed to and have stored in long-term memory; (ii) their analysis and entrenchment of the remembered forms of language into abstract categories and patterns as embodied in connected cognitive networks, and (iii) their individual agency in selecting specific forms and patterns of language to express their specific identity and intentions.

Phonology in L2 Learning

Contrasting L1 and L2 Learning

Language learning in the L1 case proceeds as a series of “baby steps” that build perceptual knowledge of the language little by little, instance by instance, through implicit learning processes, which are linked to learning about the world and build the brain’s network of associations in long-term memory. The slow, instance-by-instance build-up of implicit and perceptual knowledge about language is coordinated with the child’s developing articulatory capabilities and skills of speech production, which are gradually tuned to perception and practiced over time to match adult native speaker pronunciation standards. L2 learning, other than in

early childhood bilingualism or multilingualism, generally diverges in many ways from this course of development, occurring in different—often very different—kinds of learning contexts (e.g., school vs. home and neighborhood) and with different learning goals, priorities, and constraints that determine different learning processes and outcomes, involving transfer of skills and knowledge gained from prior experience of learning language and many other things.

Perception and Production of L2 Speech

In L1 acquisition, perception leads production to a considerable degree, and, for the first few years of life, the child is continually seeking to align speech production to speech perception.³ Once the L1 has been established, L2 acquisition proceeds in a different way, as the learner shortcuts this long-term alignment process through transfer at an early stage and may bypass natural implicit learning processes by use of learning strategies. As Strange and Shafer (2008) conclude from their review of L2 perceptual research: “It appears that extensive use of learned patterns of selective perception, in the service of robust and efficient perception of the native language, results in highly automatic patterns of perceptual processing [in the L2] that are not easily modified by subsequent linguistic experience” (p. 156) and that are linked to the motoric routines needed in speech production. Because they are highly automatic, these learned patterns of selective perception facilitate other aspects of L2 processing, while making it difficult for the learner to perceive the phonetic details and cues of the L2:

For adult learners of a foreign language, these L2 automatic selective perception routines may interfere with their ability to perceptually differentiate some phonetic contrasts in the new language. Initially, non-native phonetic segments may be perceptually assimilated to native phonological categories, resulting in perceptual confusions in tasks in which categorization is assessed. (Strange & Shafer, 2008, p. 185)

As a consequence of the perceptual magnet effect, learners will classify speech sounds in the L2 as much as possible in terms of the phoneme categories and contextual variants of their L1. Hence, L2 speech sounds

which are similar but not identical in their acoustic properties to L1 phonemes will tend to be assimilated to those L1 phonemes, by the process of **equivalence classification** that has been described in detail for L2 phonology by Flege and colleagues (Flege, 1987, 1995, 2003; Flege & Hillenbrand, 1987/1984; Flege, Monroe, & MacKay, 1995; see also Pennington, 2002). The result is **L1 transfer** and so-called “foreign accent.” As Strange and Shafer (2008) go on to remark, progress in L2 perception is possible but limited in terms of adult learners’ ultimate attainment in an L2:

[T]he ability to discriminate non-native phonetically-relevant acoustic parameters remains intact in adults and can be accessed under stimulus and task conditions that reduce cognitive demands and that allow the listeners to (learn to) attend to the appropriate acoustic structures. Thus, adult learners of an L2 can and do improve in their ability to differentiate non-native contrasts, i.e., they can develop L2 selective perception routines. However, phonetic perception of non-native contrasts may never become as automatic and robust as perception of native contrasts. (p. 185)

In terms of perception, L2 learners have to reconfigure their whole auditory space distinguishing one phoneme and its contextual variants from another, and they must learn to recognize new cues to the pronunciation of both vowels and consonants that are different from those in their mother tongue. As Couper (2015) has noted, this often means adjusting the figure-ground relationship that determines the most important, or salient, cues to a phoneme (the foreground, or “figure”) from those which are not criterial (the background, or “ground”). To do this, learners have to learn to pay attention to how speech actually sounds. This is not easy, as the focus of speech comprehension and production is most centrally on meaning, and L1 transfer provides a workable template for perceiving L2 sounds that also has entrenched, automatic connections to L1-based motor routines for speech production which make it possible to speak relatively fluently and quickly in an L2—in effect, putting pronunciation on an “auto-pilot” setting (see Chap. 1).

Language-specific “alignment of production and perception values” for different sounds as established in childhood (Flege & Eefting, 1986, p. 167) means that adult language learners’ production of L2 sounds is

strongly linked to their perception of L2 sounds (Escudero, 2005, 2007; Escudero & Boersma, 2004), both influenced by L1 knowledge. Therefore, improvement in L2 pronunciation requires modifying templates for perception and for production, shifting perception from an L1-defined acoustic space and reorienting articulation away from L1 motor templates. This is generally a gradual, incremental process, as the L2 learner's linked perception and production based on L1 transfer evolves towards the perception-production linkages of the L2. The gradual process occurs through the cumulative build-up of an internal database of L2 experience that shifts pronunciation from targets based in the learner's L1, to transitional “**interlanguage**” (Selinker, 1974/1972) targets influenced by the L1 but with an increasing strength of L2 influence over time (Major, 1987, 2001).

Progress in L2 pronunciation is generally slow and inconsistent in interlanguage development, resulting in the “**fossilization**” (Selinker, 1974/1972) or stabilization of certain error types and the phenomenon of foreign accent due to continuing L1 influence that usually persists even in very advanced learners (Long, 1990, 2015; Scovel, 2000). Global progress from *L1-based* to *L1-influenced* pronunciation can be relatively rapid in the early stages of L2 learning, as learners seek to develop pronunciation patterns according to target language models and/or the intelligibility requirements of their audience, but then slows down as learners implement their L1-influenced interlanguage pronunciation patterns automatically while they focus on other aspects of language learning. Over time, pronunciation may gradually shift towards L2 patterns, as a result of experience with the L2, but it is rare for adult language learners to ever achieve a phonological system showing no L1 influence, that is, one based entirely in L2, because of the deep entrenchment of the L1 system.

In spontaneous production, L2 learners' accented or errorful pronunciation often results from automatic application of an L1-based or L1-influenced motor template and articulatory settings (i.e., “pronunciation auto-pilot” as described previously), though it may also result from failure to properly implement an L2-based motor template, due to lack of sufficient experience or practice, or through problems of concentration or cognitive load. In more limited tasks, such as imitation or repetition in

an experimental or classroom setting, language learners' pronunciation errors may have a perceptual source, resulting from application of the wrong motor template, based on incorrect perception of what they heard according to their L1 perceptual template and retrieval of the linked L1-based production template. As in spontaneous L1 production, errorful pronunciation may not be due to perceptual error but may rather result from faulty production, that is, it may be the result of mechanical error in implementing the correct template. Although the targets for perception and production may not be modified strictly simultaneously as regards all linguistic items, forms, and types of tasks, improvement in production and improvement in perception are generally closely aligned.

Accuracy in L2 perception and production have been shown to vary by task (Bradlow, 2008; Major, 1987, 2001; Strange & Shafer, 2008; Tarone, 1988), with productive performance sometimes found to be best in focused tasks, leading perception (e.g., Flege & Eefting, 1987; Sheldon & Strange, 1982) or leading spontaneous speech (e.g., Dickerson, 1977). Yet there is considerable evidence that other than in limited contexts of training or imitation (e.g., Flege & Eefting, 1987; Sheldon & Strange, 1982), adults cannot produce phonological contrasts that they cannot perceive, so that production of L2 speech does not improve in the absence of improved perception (Escudero, 2005, 2007; Escudero & Boersma, 2004; Leather, 1999; Rochet, 1995). These results taken together suggest that pronunciation performance in production benefits by both attention and experience listening to the L2. There is also evidence that learners' most nativelike phonological performance may be found in their spontaneous speech (Hansen Edwards, 2008; Tarone, 1979, 1988). This suggests that pronunciation improves as learners are focused on communicative interaction and the generation of connected speech.

Bilinguals will often show values for L2 perception and production that are intermediate between those of monolingual speakers of their two languages (Flege & Eefting, 1987; Flege & Hillenbrand, 1987/1984; Hazan & Boulaïka, 1993). Changes that move learners' interlanguage templates to be gradually more influenced by L2 perception and production may affect their L1 perceptual and production templates over time as well, moving them towards L2 targets (Flege, 1987; Flege & Eefting, 1987; Major, 1992). However, the bulk of research (e.g., as reviewed by

Zampini, 2008, pp. 223–225), seems to show that even young bilinguals or L2 learners, those who learned a second language before the age of 6, do not have the same perceptual templates (e.g., VOT values for consonants) as monolingual speakers. Moreover, their productions may differ from those of monolingual speakers in other ways (Green, Zampini, & Magloire, 1997; Zampini & Green, 2001). The research on perception and production in both L1 and L2 speech seems to support Kuhl's (2000) view that the L1 magnet effect is established very early in language acquisition, perhaps as early as in the first year of life, so that its effects can be seen in all cases of L2 learning, at any age. The only possible exception might be children who learn two languages from birth simultaneously and so have what could be regarded as two first languages, no clear first language, or some kind of blended or creole type of first language that the infant gradually sorts into two separate languages (see Pennington, forthcoming, for discussion).

Best and Tyler (2007) suggest that for those learning an L2 after childhood, perceptual learning is largely accomplished within the first year of experience and does not improve much subsequently: “very little perceptual benefit seems to accrue from additional experience past the initial period for late learners” (p. 21). If it is true that adult language learners make little progress in perception beyond their first 12 months of experience learning an L2, this may relate to Kuhl's (2000) point about the establishment of the L1 perceptual template in the first year of L1 learning in childhood. Perhaps there is a learning process in approximately the first year of language learning, whether in L1 or L2, which establishes a perceptual basis for further learning. Yet it may not be entirely accurate to say there is **no** progress in perceptual learning beyond that point: Best and Tyler's (2007) research and a number of other studies reviewed by Flege and Eefting (1986) have shown children's perceptual learning continuing into the school years, and the literature reviewed above on L2 perception also shows the potential for continuing improvement by late learners in L2 perception over time, beyond the first year. The correct concept may be one of entrenchment or stabilization rather than fossilization or “freezing” of patterns (Ellis, 2005, p. 324), suggesting that a degree of continued learning and change is possible, after an initial period in which basic patterns are established that set limits on the degree of

further change. In both L1 and L2 learning, it can be assumed that after a certain point, new input adds less and less new information for refining the system, so that it becomes relatively stable. It may also be that in both L1 and L2 learning, the way that input is processed will be different after that point, relying increasingly on automatic processing.

Only a small amount of the language learning literature addresses matters of the social or pragmatic functions of pronunciation, and these aspects are typically not included at all in discussions of pronunciation perception. Clearly, children must focus on these aspects in order to learn them and so it would be of interest to develop a research agenda to explore how and when they are learned. It is presumably a different process from that which L2 learners go through, since they are able to transfer quite a lot of their knowledge of pronunciation pragmatics and social meaning from their L1 to the perception and production of the L2. However, that transferred knowledge is typically at best only a rough approximation to actual L2 conventions. As such, it provides general guidance for production and perception of speech but will sometimes cause misunderstanding or communication breakdown. Thus, basing L2 performance on knowledge transferred from the L1 will sometimes cause a language learner's production of a message to be distorted, so that listeners do not interpret it in the meaning which the learner intended to convey, and will also sometimes cause a learner's perception of a message to be distorted, so that the learner does not interpret it in the meaning which a speaker intended to convey.

Factors Affecting L2 Pronunciation

There is a considerable amount of evidence that language learners beyond a certain age have difficulty achieving full native speaker proficiency in pronunciation (Flege, Yeni-Komshian, & Liu, 1999; Graena & Long, 2013a, 2013b; Long, 1990, 2015; Scovel, 2000). As Long (2015) summarizes, “the best results for adults are poorer than the routine achievements of young children with early sustained exposure, even when conditions are seemingly optimal, i.e., when learner profiles (attitude, motivation, etc.) are positive, time to learn is unlimited, and usable input

is plentiful” (p. 38). Yet, as reviewed by Saito (2015b), there is also a considerable amount of evidence that language learners at any stage of life can improve their pronunciation over time, and “some learners are able to attain high-level L2 performance” (p. 741). The following factors have been shown to be influential in L2 learners’ ultimate attainment and can specifically impact L2 pronunciation:

- Transfer and other learning processes
- Age effects
- Quantity and quality of input and output
- Educational factors
- Individual differences

Transfer and Other Learning Processes

A central concept in L2 learning is that of transfer of prior learning to subsequent learning, specifically, L1 transfer, or the learning of a second language on the basis of the mother tongue. According to Lado (1957), a methodology of systematic comparison of the native language to the language that is the target of learning by means of **Contrastive Analysis** (CA) would indicate the degree of ease or difficulty of learning a second language. In practice, this assumption, known as the **Contrastive Analysis Hypothesis** (CA), has been tested by sampling learners’ speech to note which aspects were produced correctly and which were produced incorrectly according to native speaker norms, then looking to the learner’s L1 to determine if its similarity or difference from the L2 could explain the specific pattern of correct and incorrect forms, considering both **positive transfer**, or the possible facilitating influence of the L1, in the case of correct performance, and **negative transfer**, or **interference** from the L1, in the case of incorrect performance.

Early studies of L2 phonology from a CA perspective (e.g., Brière, 1966, 1968; Johansson, 1973; Nemser, 1971; Stockwell & Bowen, 1965) discovered that L1–L2 similarity/difference could explain a portion of L2 learners’ performance but in general the CA was not strongly supported. In fact, it seems that ease versus difficulty of L2 learning and correct versus

incorrect L2 pronunciation are not related to L1–L2 similarity and difference in any direct or obvious way, since features of greatest difference between languages are often not those which cause the most difficulty for learners or the most serious or frequent errors. Rather, the most common and persistent error types often have to do with the most similar sounds across two languages. Wode (1977) stressed cross-language similarity as the key issue in L2 transfer and pronunciation difficulties, and he further suggested (Wode, 1983) that when L1/L2 phones were not sufficiently similar, transfer would not operate and learners would apply other pronunciation strategies, resulting in error patterns similar to those of L1 acquisition. The fact that L1 transfer (and equivalence classification) does not occur without perceived similarity in L1 and L2 forms is an important fact for language teaching, highlighting the potential value of focusing on differences between L1 and L2 forms (e.g., in form-focused instruction; see Chaps. 3 and 4).

Besides those due to L1 interference, L2 errors were also seen to involve learning processes that occur in child language acquisition, such as simplification, approximation, and (over-)generalization of forms and rules (Richards, 1974/1971). Since these types of errors are common in L1 acquisition, they were considered to be the result of natural language acquisition processes favoring language universals such as unmarked (simple) over marked (complex) forms. Within the construct of interlanguage as a learner's evolving linguistic system bridging between the L1 and the L2 (Selinker, 1974/1972), researchers in L2 phonology aimed to determine which of learners' errors that could not be attributed to L1 transfer could be described as **developmental** errors (Hecht & Mulford, 1987/1980; Tarone, 1987/1978) or as evidence for natural phonological processes or language universals (Altenberg & Vago, 1987/1983; Eckman, 1987/1977, 1987/1981). Eckman (1987/1977) sought to supplement the CA with the **Markedness Differential Hypothesis** (MDH), proposing that, regardless of L1–L2 similarity or difference, unmarked (simpler or more basic) linguistic features would be acquired before marked, which are more complex and difficult to produce.

Combinations of transfer, developmental processes, and “rapid speech phenomena” were posited by Hecht and Mulford (1987/1980) as sources of the L2 phonology of an L1 Icelandic-speaking child learning English

after coming to the United States at age 6. For adult learners, phonological transfer and a number of other factors have been noted as contributing to L2 phonology. Altenberg and Vago (1987/1983), in their study of the errors made by adult L1 Hungarian learners of English as L2 when reading aloud from a passage of continuous text, noted pronunciation based on sound-spelling correspondences in the L1, pronunciation of a morphologically related word (e.g., *width* pronounced as *wide*, *considerably* as *considerable*), and possible mechanical difficulties of articulation or psychological difficulties having to do with hesitation and uncertainty regarding the pronunciation of specific words (Altenberg & Vago, 1987/1983). Knowledge of lexis and of the written forms of words are obvious factors in these cases, indicating complex sources of errors combining L1 transfer, L2 spoken and written language knowledge, and task constraints. Beebe (1987/1984) also noted productions of English consonants by L1 Thai and Japanese learners that were sequential composites or mergers of an L1 and L2 sound, such as [θs] or [sθ] for English /s/ and [rl], [lr], or [wl] for English /l/. These examples suggest some of the difficulties involved in classifying learner errors as resulting from L1 transfer. Other complex cases arising from analyses of L2 phonological errors seem to indicate learners' transfer of L1 stylistic variants to L2 communicative situations (Beebe, 1987/1980, 1987/1984; Schmidt, 1987/1977) and their variable performance in different contexts, some of which show systematic variation and style-shifting (Adjémian, 1976; Dickerson, 1975; Tarone, 1979; Weinberger, 1987; Zuengler, 1982, 1988).

Age Effects

Age effects in L2 learning may involve continuous cognitive aging (Birdsong, 2005, 2006; Hakuta, Bialystok, & Wiley, 2003) as well as maturational constraints or sensitive periods for language acquisition (Granena & Long, 2013a, 2013b; Long, 1990), both of which can affect perception as well as production. Many studies have shown that age of arrival (or "age of acquisition"; Saito, 2015b, p. 714) is a strong predictor of ultimate attainment in a second language, notably in pronunciation, though the effects do not necessarily apply for learners past puberty

(Abrahamsson, 2012; DeKeyser, Alfi-Shabta, & Ravid, 2010; Flege, Munro, & MacKay, 1995; Flege et al., 1999; Granena & Long, 2013a, 2013b; Hopp & Schmid, 2013; Johnson & Newport, 1989; Oyama, 1976; Patkowski, 1990; Scovel, 2000). Thus, when it comes to learning a language, “the younger, the better.” Bley-Vroman (1988, p. 22) mentions a study showing a correlation of $-.7$ between the age of arrival of immigrants to the United States and their ultimate attainment, meaning that there was a 70% connection between a younger age and language proficiency for immigrants learning English as L2.

The strong statistical relationship between age of arrival or age of acquisition and ultimate attainment in L2, especially in pronunciation, may be evidence of a cognitively defined “critical period” for language acquisition, after which language learning will be much less effective. According to the **Critical Period Hypothesis**, age effects in language learning are due to permanent changes in the brain that occur in childhood and create “a discontinuity in learning outcomes that corresponds to a maturational point in...development.... The discontinuity can be established either by the cessation of learning (strong CP) or a change in the slope of the learning curve (weak CP) after the close of the critical period” (Bialystok, 2002, p. 482). The exact point of discontinuity, the end-point of the critical period, is debated, with some putting it as early as age 5 or 6 and others as late as age 12, 13, or even 16. Others argue for a maturationally determined series of “sensitive periods” for language acquisition ending at puberty or somewhat later in adolescence (Graena & Long, 2013b).

The fact that a small minority of individuals appear to learn a second or additional language well, with little or no discernible foreign accent even when beginning language learning in their teen years (Bongaerts, 1999), would appear to provide counterevidence to a critical period for language acquisition. Those who support the existence of a critical period in language learning argue, however, that the reason some post-pubertal language learners are able to improve in their L2 at a late age, especially their pronunciation, is because of individual differences in language learning abilities and circumstances that are not relevant in L1 acquisition.

An alternative explanation for age effects in older learners is what Saito (2015b) labels the “**Cognitive Aging Hypothesis**” (CAH), according to

which there is no specific cut-off point to language acquisition because language learning capabilities can still be active at any age (e.g., Bialystok, 1997, 2002; Bialystok, Craik, Klein, & Viswanathan, 2004; Bialystok & Miller, 1999; Birdsong, 2005, 2006; Derwing & Munro, 2013; Flege, 2003, 2009; Hakuta et al., 2003; Hopp & Schmid, 2013), assuming psychological and social conditions of learning that are similar to those of young learners, including high access to and use of the L2. In addition, older learners may benefit in language learning from their greater knowledge and experience of learning and of life in general. At the same time, there are biological declines in cognitive functions which are part of the aging process and which may make some aspects of language learning (as well as other kinds of learning) less effective, so that age of arrival effects are still in evidence in older learners, noticeably, for accentness (Birdsong, 2005, 2006; Hakuta et al., 2003). Thus, older learners may have advantages related to prior knowledge and experience in vocabulary and grammar, and they may also have disadvantages related to prior knowledge and experience in pronunciation (Derwing & Munro, 2009; Derwing, Rossiter, Munro, & Thomson, 2004; Saito, 2015a, 2015b).

The already fully formed L1 of late language learners inevitably affects their L2, notably in pronunciation. At every stage of L2 learning, learners' phoneme targets will be set at some distance from those of the L2 and will show the influence of L1 targets. In addition, continuing exposure to and use of an L2 affects their L1 performance (Bialystok & Miller, 1999; Flege, 2007, 2009; Flege & Eefting, 1986, 1987; Flege, Frieda, & Nozawa, 1997; Hopp & Schmid, 2013), as the auditory and articulatory positions for vowels and consonants move away from those of the L1 and towards those for L2. Thus, learners' phoneme targets, whether speaking L1 or L2, are intermediate between those for the two languages. As Birdsong (2014) points out, the fact that bilinguals show "non-monolingual nativelikeness in the L1 is suggestive of a capacity to learn language in adulthood" (p. 46): "Arguably, the fact that the L1 can be influenced by the L2 in adulthood is evidence for maturationally conditioned representational plasticity" (*ibid.*)—in other words, the opposite of a critical period entrenching the L1. Moreover, "[s]ince bilinguals are not like monolinguals in either of their languages, it is hard to argue that comprehensive

nativelikeness, scrutinized or not, should be held up as the gold standard for falsifying the [Critical Period Hypothesis for L2 acquisition]” (Birdsong, 2014, p. 47).

Quantity and Type of Input and Output

L2 pronunciation is affected by the quantity of L2 input and use (Flege, 2009; Flege et al., 1995, 1997, 1999; Moyer, 2004; Piske & MacKay, 1999) and by the quality or type of L2 input in classroom or natural communicative settings (Long, 1981, 1983, 1996, 2015), which may vary greatly depending on the extent to which the learner is immersed in an L2 rather than an L1 environment. Given the feasibility of the Cognitive Aging Hypothesis, quantity and quality of input and output provide alternative explanations for the strong statistical relationship between age of arrival or age of acquisition and ultimate attainment in L2, especially in pronunciation.

Arriving in an L2 environment at an early age tends to mean becoming immersed in L2 speech in the mainstream community, particularly, through early schooling, so that age of arrival may be at least in part a proxy measure for amount and/or quality of exposure to and/or interaction in the L2.⁴ Research by Flege (2009) suggests that relating degree of foreign accent to length of residence or age of arrival may mask the key factor of amount of L2 input which long-term residents in an L1 environment have been exposed to. From his analyses, it seems that degree of foreign accent may be associated directly with frequency of use of the L1 and nonuse of the L2. As contrasted with early arrivals, later arrivals may continue to use their L1 in preference to their L2, even after many years in a country where the L2 predominates. However, it must be noted that there has been very little study by L2 researchers of amount of language use other than indirectly, that is, by self-reported estimates by L2 speakers.

The quality or type of input to L2 learners may be restricted to only certain kinds of exposure or interaction depending on the learner’s need for the L2 and social situation in relation to L2 speakers, as social factors may limit L2 learners’ “access to L2 use and linguistic environment”

(Hansen Edwards, 2008, p. 272). In a classroom situation of foreign language learning, both the quantity and quality of input are reduced, often very considerably, as compared to learning a language naturally in a wide range of communicative tasks and contexts. The manner of learning in school and non-school contexts (e.g., through reading and study vs. through involvement in tasks requiring listening comprehension and speaking) is another factor that affects learning process, quantity, and quality of what is learned.

Educational Factors

Level of education or years of schooling affect amount of learning and also the efficiency and effectiveness of learning through application of learning skills, so that L2 learners with higher education levels often achieve a higher level of language proficiency (Birdsong, 2014; Derwing & Munro, 2009; Spada & Tomita, 2010). The type of education received in an L2 affects what is learned, as in all kinds of instructed learning, and language skills in a learner's L1 are also related to those in the L2 (Miettinen, 2012; Rimfield, Dale, & Plomin, 2015).

Literacy and an understanding of words and their parts play an important role in phonological processing and in explicit learning in both L1 and L2 (Reis & Castro-Caldas, 1997; Tarone, Bigelow, & Hansen, 2009). While implicit or automatic phonological processing is an aspect of the amassing of a linguistic database and the statistical processing of linguistic sequences in the brain that Ellis (1994, 1996, 2001, 2002) has described, it appears that explicit or controlled phonological processing beyond the first years of language learning is largely dependent on the visual form of words learned through reading and writing (Reis & Castro-Caldas, 1997). According to Reis and Castro-Caldas (1997), nonliterate children and adults are dependent on semantic processing as a largely automatized default strategy for creating meaningful language, whereas literate individuals can use explicit phonological processing of oral language involving a focus on form.

In the L2 case, the default semantic processing strategy, combined with interference or transfer from the native language, affects phonological

processing, shortcircuiting and biasing both implicit or automatic processing and explicit or nonautomatic processing by equating L2 phones with L1 phones, based on only a rough equivalence in their phonological form that leads to fossilization or stabilization of pronunciation in a foreign accent. For literate adults, the equivalence classification operates in part via the sound-spelling correspondences of the mother tongue. Only through long exposure and attention to salient instances of input that highlight the unique features of L2 forms will L2 learners progress in their pronunciation. This is the basis of focus-on-form initiatives in instruction (for review, see Wulff & Ellis, 2018; see also Chaps. 3 and 4). However, the ability to benefit by focus-on-form input may depend on literacy.

Research by Tarone et al. (2009) reveals that low-literate adolescent and adult L2 learners have “significant difficulty completing oral tasks that require the noticing and manipulation of linguistic form” (p. 73). Tarone et al. (2009) show that nonliterate late language learners have a strong focus on meaning, relying on a semantic processing strategy in their language production, and cannot easily focus on form, even in imitation tasks (*ibid.*, chap. 5) and even after multiple recasts have been provided (*ibid.*, chap. 4). They give evidence that a nonliterate speaker is attentive to salient cues in those recasts, such as contrastive stress, since he repeats it in his response to the feedback given, but places it on the wrong word. As they conclude, “The ability to attend to and analyze oral L2 input in terms of segmental linguistic units may depend on an individual’s prior alphabetic print literacy level” (Tarone et al., 2009, p. 73).

Individual Differences

Beyond the learner’s L1, age, language input/output, and educational factors as reviewed above, individual differences affect the ultimate attainment of learners in their L2, and specifically in pronunciation. Dörnyei (2006) identifies the five most important individual differences as personality, aptitude, motivation, learning styles, and learning strategies. We note that what have been called learning styles are sometimes referred to as “cognitive styles” and may not be clearly distinguishable from personality factors and also that the attributes variously termed personality, cognitive

styles, and learning styles are increasingly being developed into models whose core construct is motivation. The whole notion of individual difference factors in language learning is in flux, as researchers have increasingly realized how important they are in explaining language learning outcomes, once the focus is taken off innate factors of Universal Grammar and cognitive maturation. The trend of the research on individual differences is to highlight the very significant role of all kinds of individual differences on language learning outcomes and the fact that these individual differences are found in both L1 and L2 learning.

Aptitude

Some people seem to have special talent or ability for learning languages that has a large inborn component, though it may also show the effects of learning and experience (Sáfár & Kormos, 2008). Language learning ability may overlap general intellectual skills or cognitive abilities, including memory (Skehan, 1998), and is related to musical ability (Hu et al., 2013, p. 367), but can nonetheless be measured separately as a person's **language aptitude**. Differences in language aptitude are a key factor predicting ultimate attainment in an L2 (Long, 2015, pp. 58–60), and those differences have been claimed to account for more of the variance in L2 performance than other individual factors (Ehrman & Oxford, 1995; Rimfield et al., 2015).

Language aptitude includes special ability or talent in pronunciation, which may be associated with musical ability (Baran-Lucarz, 2012a). Hu et al. (2013) observe that “adults vary greatly in their L2 pronunciation aptitude—both with respect to segmental (speech sounds) and suprasegmental (intonation, rhythm etc.) manifestations of spoken language” (p. 366). Purcell and Suter (1980) demonstrated that the ability to imitate sounds in a foreign language was a significant predictor of L2 pronunciation accuracy. Carroll (1981), one of the developers of an early test, the **Modern Language Aptitude Test** (MLAT; Carroll & Sapon, 2002/1959), recognized phonetic sensitivity or **phonetic coding ability** (the ability to discriminate different sounds and sound-symbol correspondences and to retain these in memory) as one of four components of

language aptitude assessed by the MLAT, and the LLAMA Language Aptitude Test (Meara, 2005), like the MLAT, includes subtests (LLAMA D and E) involving these abilities (see Chap. 6). Phonetic sensitivity or coding ability seems to be differentially represented in the human population as an inborn trait, though it also may be affected by L1 literacy.

Skehan (1998, chap. 9) reviewed studies showing that “very weak foreign language learners” are weak on language aptitude measures of phonemic coding ability, which he describes as their lacking “input skills,” whereas “[e]xceptionally successful foreign language learners consistently seem to be characterized by the possession of unusual memories, particularly for the retention of verbal material” (p. 233). In his view, phonemic coding ability relates to L2 proficiency at a relatively low level of language aptitude while memory is the key proficiency variable at a high level of language aptitude (Skehan, 1998, pp. 217–218). Granena and Long (2013a) found that older learners’ assessed degree of foreign accent is linked to their language learning aptitude as measured by the LLAMA E scale for recognizing correspondences between sounds and symbols and the LLAMA F scale for inferring grammatical rules in an unknown language. Saito (2017) found that LLAMA E scores of Japanese L1 learners of English as a foreign language were predictive of their accuracy in pronunciation and grammatical morphology. Such data are fit into a critical period explanation by claiming that critical period effects are relevant only to the automatic and implicit aspects of language learning which are hypothesized to be the same for all.

Personality and Cognitive/Learning Style

Relatively stable individual traits that have been shown to be related to language ability or achievement and motivation include those classified as personality characteristics and other traits classified as personality or cognitive style, or as cognitive style or learning style. These include:

- the “**Big Five**” **personality dimensions** of extraversion, neuroticism, conscientiousness, agreeableness, and openness to new experiences (Pervin & Cervone, 2010), and possibly also the **Multicultural**

- Personality Questionnaire** (Van Der Zee, van Oudenhoven, Ponterotto, & Fietzer, 2013) dimensions of cultural empathy, open-mindedness, social initiative, emotional stability, and flexibility;
- other traits differentially classified as personality or cognitive style—tolerance of ambiguity, empathy, risk-taking, self-esteem (global, situational, task)—or as cognitive style or learning style—field independence/dependence, reflectivity/impulsivity (Brown, 1994, chap. 6; Larsen-Freeman & Long, 1991, pp. 184–196).

Personality and cognitive style or learning style affect the way people engage with other people and with information and so influence the type and amount of input, interaction, and output which learners will involve themselves in (Brown, 1994, chap. 6; Dewaele, 2013; Schmidt, 1983; Schmidt, Boraie, & Kassabgy, 1996; Schmidt & Watanabe, 2001) and also their manner of processing input (Hu et al., 2013; Reiterer et al., 2011; Skehan, 1998, chap. 9). Personality and cognitive style or learning style are interactive with motivation and often discussed within considerations of motivation, as we will do as well. Our review of literature is selective, focusing on results that have to do specifically with pronunciation or related skills.

Extraversion

Extraversion is defined as directing attention toward and obtaining gratification from external stimuli, involving especially responses and reinforcement from other people, while introversion is defined as directing attention toward and obtaining gratification from internal stimuli, involving especially one's own thoughts and mental world. An extravert tends to be talkative and sociable, whereas an introvert tends to be self-contained and reserved. Although some early studies (e.g., Naiman, Fröhlich, & Stern, 1978) found no affect for extraversion/introversion on language performance, Dewaele and Furnham (2000) make a strong case “that extraversion is inextricably linked with fluency in second language (L2) production” (p. 356), citing research showing that “extraverts have a better short-term memory; are more stress-resistant and are less anxious in second language production” (ibid.)—all traits which aid in their production of L2 speech.

Their superior short-term memory seems to aid their L2 production as they are better able to handle the multiple cognitive and mechanical demands of sequential speech production, and their lower stress and anxiety avoids additional short-term memory distractions and helps keep their attention focused on speech production.

As an important factor contributing to fluency, the extraverts in Dewaele and Furnham's (2000) study of L1 Flemish university student advanced learners of L2 French produced speech at a much faster rate than the introverts. Whereas "extraverts..., being better equipped to cope with interpersonal stress, are able to maintain most of their automatised processing" (ibid.), introverts may be

...unable to maintain the same level of automaticity of speech production when they are under some sort of arousal/stress (being observed or tested). They slide back to controlled processing which overloads their working memory. This means their speech slows down, they hesitate more often, they tend to make more errors and they are unable to produce utterances of great length. (Dewaele & Furnham, 2000, pp. 362–363)

The researchers note an overall difference in the speaking style of the extraverts versus the introverts, which they characterize as an implicit versus and explicit style. The implicit style assumes "a shared spatio-temporal context" of speaker and hearer while an explicit style creates that context within the speech itself, through "explicit and precise description" (Dewaele & Furnham, 2000, p. 360), as a way to ensure that the speaker will not be misunderstood.⁵ The explicit style of speech requires more time and therefore more work by short-term memory to access low-frequency words (Dewaele & Furnham, 2000, p. 363), which is an important factor in slowing down the introverts' speech production and so impacting their fluency.

Neuroticism and Anxiety

Neuroticism is a Big Five personality trait defined by high anxiety. Language learning in a classroom setting can create a specific form of negative anxiety, "associated with an arousal of the autonomic nervous system," which has

been termed **Foreign Language Classroom Anxiety** (FLCA; Horwitz, Horwitz, & Cope, 1986, p. 125). FLCA is often directly related to a fear of speaking, which may be greater in a classroom or in a formal or public situation, which may affect certain types of learners (e.g., introverts, Dewaele & Furnham, 2000) more than others, and which may be expected to have effects on pronunciation. Dewaele (2013) found a link between neuroticism and FLCA in learner groups at universities in Spain and the United Kingdom. He also found that FLCA showed strong correlations across the individual languages participants knew, suggesting that it is a relatively stable individual difference which is not greatly affected by context.

Further research by Dewaele and Al-Saraj (2015) with Arabic learners of English revealed that FLCA was negatively related to dimensions of the Multicultural Personality Questionnaire, suggesting that aspects of a multicultural personality orientation are associated with a non-anxious mindset for language learning. In addition, the researchers found that those with high self-reported use of English and proficiency in the L2 were also those with low FLCA. Learners' self-assessed L2 proficiency had the strongest value as a predictor, suggesting that confidence and a sense of self-efficacy are related to maintaining a positive mindset in the language learning classroom that might make a student more willing to speak and thus give the best chance for developing pronunciation competence.

Baran-Lucarz (2014) has sought to isolate the aspect of foreign language classroom anxiety that is specifically related to pronunciation, developing a construct of **pronunciation anxiety** (PA) that includes the factors of pronunciation self-perception, fear of negative evaluation, and beliefs concerning the pronunciation of the target language. In her research with Polish EFL classroom learners, she found that PA was related to level of familiarity with interlocutor(s), group size, type of task, and target-language proficiency level. Pronunciation anxiety was also negatively related to participants' assessed willingness to communicate (see below).

Tolerance of Ambiguity

There is evidence that **tolerance of ambiguity**—the willingness to accept uncertainty and not feel a need to rush to a decision or closure—is related to L2 performance. In early studies, Naiman et al. (1978) reported significant

correlations between tolerance of ambiguity and measures of listening comprehension but not sentence imitation while Chappelle and Roberts (1986) reported correlations between measured tolerance of ambiguity and end-of-course L2 proficiency. More recently, Baran-Lucarz (2012b) discovered a weak correlation between scores for accuracy in L2 English pronunciation and L2 speakers' "accept[ance of] objects, concepts and situations that lack clear borders" (p. 60), which Dewaele and Li Wei (2013) link to tolerance of ambiguity (p. 233). In their view, "A moderate level of [tolerance of ambiguity]...seems optimal in SLA" (ibid.). Specifically for pronunciation, we suggest that tolerance of ambiguity might mean not immediately jumping to a transfer or equivalence classification strategy for L2 perception and instead being willing to spend the time and to give the attention needed to carefully observe and sort out perceptual and articulatory differences between the two languages. Dewaele and Li Wei (2013) point to research by Reiterer, Singh, and Winkler (2012) showing that high ability language imitators were also high in measures of articulatory flexibility and "must still possess this openness to build new phonetic categories on an ad-hoc basis, and not rely on pre-experienced, entrenched categories" (p. 16). Dewaele and Li Wei (2013) observe the similarity of this notion of articulatory flexibility to their definition of tolerance of ambiguity, "namely the capacity to perceive and process information that deviates from the usual patterns" (p. 232).

Tolerance of ambiguity would seem to be related to the personality or cognitive style characteristic of **reflectivity** that has been related to successful language learning (Brown, 1994, chap. 6) and possibly also to the personality dimensions of **agreeableness** and **openness to new experiences** which have been described in another area of research as related to empathy (e.g., Djikic, Oatley, & Carland, 2012) and in motivation studies as related to integrativeness (Dörnyei, 2006; Gardner, 2007; see below).

Empathy

Empathy is being willing and able to take the perspective of another and showing sensitivity to their circumstances. Guiora, Lane, and Bosworth (1967) reported significant correlations between empathy and pronunciation accuracy for teachers of French. Guiora (1972)

posited the construct of “**language ego**,” which he related to empathy, arguing that the more permeable a person’s “ego boundary” (empathic capacity) is, the better the person’s pronunciation. Guiora and colleagues studied ego permeability or empathy as an aspect of pronunciation and fluency in experiments with alcohol, which in small amounts had a facilitating effect on pronunciation (Guiora, Beit-Hallahmi, Brannon, Dull, & Scovel, 1972), and relaxant drugs such as Valium, which did not improve L2 pronunciation (Guiora, Acton, Erard, & Strickland, 1980). Hu et al. (2013) investigated a group of advanced L1 German learners of English, assessing their English pronunciation aptitude (by reading a passage aloud) and phonetic coding ability (by a subtest of the MLAT involving knowledge of sound-spelling correspondences) along with intelligence, musical aptitude, phonological working memory (by number sequence and nonsense word repetition tasks), and personality. The English pronunciation score correlated significantly with music aptitude, personality dimensions of openness and empathy, and—most strongly—phonetic coding ability. A regression analysis “showed that phonetic coding ability and empathy, but not phonological working memory” (p. 366, Abstract), predicted the Germans’ L2 English pronunciation aptitude. This is not surprising given that pronunciation aptitude and coding ability were both tested in ways related to knowledge of literacy and written words.

Field Independence

The cognitive/learning style variable of **field independence** (FI) has received considerable attention in the SLA literature (Larsen-Freeman & Long, 1991, p. 193). Field Independence is a perceptual-analytic ability assessed by the Group Embedded Figures Test (Witkin, Oltman, Raskin, & Karp, 1971) to determine the ability to distinguish a simple shape or picture that has been embedded in a more complex shape or picture. The more complex shape or picture is called the field, so people who can find the simple shape or picture within this more complex background field are called **field independent**. People who cannot do this easily are called **field dependent**. In general, field independence is the ability to focus on

and separate details from a larger context. Field independent learners tend to have an analytical orientation to perception and learning whereas field dependent learners tend to have a holistic orientation to perception and learning.

This cognitive style dimension has been found to relate to a number of other individual differences (Witkin & Goodenough, 1981), including school achievement and scores on standardized tests, people's chosen fields of study, and general differences in national cultures (Hansen, 1984). In their characterization of the good language learner, Naiman et al. (1978) observed a relationship between FI and proficiency in listening and sentence imitation, and Hansen and Stansfield (1981) and Chapelle and Roberts (1986) reported correlations between FI and tests of language proficiency.

Skehan (1998) suggested that field independence can be related to input-processing:

In the case of auditory material, the learner has to extract what is important from the stream of incoming sound. This sound will contain a great deal of irrelevant information, and it will be advantageous to devote attention to features which help meaning to be recovered. In this respect, field independence would relate to...attentional capacities..., with [field independent] individuals having greater capacity to channel attention selectively and notice important aspects of language. (p. 238)

Field independent learners therefore might have an advantage in the ability to focus on form and to notice cues to L2 differences in form and meaning that field dependent learners might miss. A recent study of field independence found that language learners high in this trait significantly outperformed other learners in their ability to benefit from recasts (Rassaei, 2015), suggesting that field independence may aid learners to benefit from **form-focused instruction (FFI)**. Recent work by Baran-Lucarz (2012a) determined that field independence predicted high-level performance in pronunciation, as did a preference for auditory learning. We note that these two characteristics together might result in an auditory focus on form that should be facilitative for pronunciation.

Identity

In the L1 case, the motivational basis for language learning is social, “emanat[ing] from (1) a desire to communicate with other persons, and (2) a desire to be like other persons (that is, to imitate them)” (Tomasello, 2003, p. 31). Thus, language learning in childhood is tied in with developing a social identity. L1 learning is also tied in with developing one’s individual identity in terms of vocal characteristics and personality. This is true to a lesser extent in the L2 case, in which learners already have a core identity connected to their personal characteristics and to their native language and culture. They may nonetheless modify aspects of their identity or evolve new identity facets in relation to their social goals and interactions with L2 speakers:

Since a language is a communicative vehicle and also a repository of culture, learning a new language makes it possible to expand oneself by developing a new communicative and cultural repertoire.... At the same time, language learners carry along aspects of their core cultural and ethnic identity that may provide valuable stabilizing influences but may also cause them to resist referencing their identity to new linguistic practices and their associated discourses, cultural attributes, and communities. (Pennington, 2018, p. 94)

According to Hansen Edwards (2008), social factors that impact “learners’ abilities to gain access to L2 use opportunities and the density of this access, as well as attitudes to the L1 and L2 community...may affect not only the learners’ use of L2 but also their perceptions of their own L1 and L2 identities, and therefore, their willingness—or lack thereof—to acquire and/or use the appropriate speech markers to signal belongingness in that community...” (Hansen Edwards, 2008, p. 273). Learners may intentionally diverge from standard language patterns for certain reasons. Research has shown that

...learners are active agents in choosing not only what and how they use their L2, but also in choosing the L2 target, and therefore what they acquire of the L2.... Much of the research on social factors...has shown that learners

are sophisticated L2 users and L2 learners, and they are active agents in what elements of the L2 they target for acquisition and/or use in different contexts. For example, research has indicated that learners are able to accommodate their speech to their interlocutor based on perceived similarities such as ethnic identification...and occupation, education, and gender.... Additionally, learners may be aware of how certain variants are used by speaker in different context/communities. Therefore, they may actively use (or avoid using) some variants or linguistic features over others based on gender, ethnic, national identities... and peer group identifications.... (Hansen Edwards, 2008, p. 272)

Group (ethnic) identity is an important social factor in language that affects language learning and performance (Gatbonton, Trofimovich, & Segalowitz, 2011; Gumperz, 1982; Sachdev & Giles, 2004; Schumann, 1978) and, specifically, the adoption of pronunciation features (Fought, 2002, 2006; Giles, 1979; Zuengler, 1988). Pronunciation may in fact be a stronger feature of group identity than other aspects of language, as suggested by the fact that minority groups may not adopt features of majority group pronunciation as readily as its grammatical features (Fought, 2002, pp. 449–450). As Giles (1979) observed, specific features of L1 pronunciation, rather than representing “interlingual *interferences*, particularly in the cases of second and third generations of immigrants,... may often be adopted by them deliberately as ethnic speech markers to establish a distinctive linguistic identity” (p. 260, emphasis in original). In childhood bilingualism, arrival in a speech community at an early age opens the possibility of a strong connection of the L2 to the child’s developing identity and group affiliations as referenced by accent, whereas an adult’s identity and group affiliations, together with identifying accent and other linguistic features, will be relatively entrenched and therefore resistant to change.

Pennington, Lau, and Sachdev (2011) found differential adoption of London sociolinguistic variables by second-generation Chinese and Bangladeshi adolescents in London that may indicate a difference in desired model for behavior or in exposure to input, both of which would seem to be related to frequency of association with groups modeling those variables. For example, the incidence of high rising tone (HRT) or

“Upspeak” (Bradford, 1997) in statements, a sociophonetic variable that is widespread in the English-speaking world among what Cruttenden (1997/1986) calls “New Yuppies” (p. 130), was high in both interview and peer conversations among the Chinese youth whereas the Bangladeshi participants had only three instances of this variable in interviews and none in peer conversations. The differential pattern of adoption of this and other variables by the Chinese and Bangladeshi youth is consistent with different sociolinguistic identifications and patterns of interaction in the community.

In another study, aspects of Québec Francophones’ measured ethnic group affiliation have been shown to be negatively associated with five global measures of oral proficiency—native speaker ratings of fluency, accentedness, and comprehensibility, and both native speaker and self-ratings of global proficiency (Gatbonton & Trofimovich, 2008)—and also with their accuracy in pronouncing the English voiced interdental fricative /ð/, known to be an important marker of ethnolinguistic identity (Gatbonton et al., 2011). However, the associations of the measures of ethnic group affiliation and of oral proficiency “vanished when [participants’] self-reported amount of L2 use was partialled out” (Gatbonton et al., 2011, p. 198). Both the lack of input and the desire not to imitate the behavior of speakers outside one’s ethnic group may be causative factors in L2 speakers’ lack of production of particular phonological features.

Motivation and Affective Factors

Motivation is whatever propels a person to do something, that which both initiates and sustains activity and effort. Gardner (2007) describes motivation as “a multifaceted construct” with a number of attributes and sources:

The motivated individual is goal directed, expends effort, is persistent, is attentive, has desires (wants), exhibits positive affect, is aroused, has expectancies, demonstrates self-confidence (self-efficacy), and has reasons (motives). All of these attributes characterize the individual who is motivated to learn a language. (p. 15)

A main source of motivation is the learner's need or reason for learning the language. In Gardner's (1982, 1985) original conception, motivation was classified as either **integrative motivation** (i.e., reflecting a desire to integrate or affiliate with native speakers of the L2) or **instrumental motivation** (i.e., reflecting practical or utilitarian reasons for learning the L2). A different division has been made by Deci and Ryan (1985) into two types of motivation, **intrinsic motivation** (i.e., reflecting internal needs, desires, and satisfactions) and **extrinsic motivation** (i.e., reflecting external considerations and constraints). As applied to language learning, intrinsic motivation might embody a love of the language or of the language learning process, whereas extrinsic motivation might embody the expectations of one's family or employer or the need to pass a language test for graduation.

Intrinsically motivated students may have less language anxiety and a greater sense of self-efficacy and agency than extrinsically or instrumentally oriented students (Richards, 1996), and this may help to maintain learning effort and satisfaction at a high level. Those with an integrative or intrinsic motivation may have a greater focus on pronunciation, or a greater desire to speak without an L2-influenced accent, than those with an instrumental or extrinsic motivation (Sardegna, Lee, & Kusey, 2014; Smit, 2002), and this may make them good pronunciation learners, based on their strong internal motivation (Brown, 2008). In this connection, Szyszka (2015) found that a group of learners who were highly proficient in L2 pronunciation rated their concern for L2 pronunciation as the most important factor affecting their L2 pronunciation.

Gardner (2007) has modified his motivational terminology, favoring the term "openness" or "openness to cultural identification" for what he had originally classified as integrative motivation:

In our original research we labelled this component Integrativeness and focussed attention on the individual being interested in learning the language in order to interact with valued members of the other community and/or to learn more about that community (i.e., an integrative orientation and favourable attitudes toward the community), but in later research we found that it could also involve an open interest in other cultural

communities in general (i.e., an absence of Ethnocentrism and authoritarianism, or the presence of Xenophilic attitudes, etc.), which we measured with our Interest in Foreign Languages scale. (p. 15)

Based on his research, Gardner (2007) maintains that the most influential variable on motivation in L2 learning is integrativeness/openness and the second-most influential one is attitudes towards the learning situation. Researchers in social psychology have noted the social aspect of motivation that underlies the desire for integration or affiliation as well as the desire to maintain a distinctive identity or a degree of non-affiliation (Clément & Noels, 1992; Giles & Byrne, 1982; Hutnik, 1991; Sachdev & Bourhis, 2005; Sachdev & Giles, 2004), which can be considered the negative side of the integrativeness/openness construct.

Dörnyei (2006, pp. 52–53) observes that the traditional notion of integrative motivation as assimilation into an L1 native speaker community does not necessarily apply in an EIL context, where the learner's goal may be to be part of a global community of English speakers or to develop a Global English identity. A goal of international identity or global integration is obviously different from the traditional goal of integrative motivation built on a native speaker model for performance. As Dörnyei (2006) notes, it blends aspects of integrative and instrumental motivations—a desire to be a citizen of the world as a transnational English-speaking community and also a desire to have full access to technological advances, which a knowledge of English facilitates. We note that a goal of developing a Global English identity might in this sense set a somewhat higher or different bar for performance than a goal of acquiring **English as a lingua franca**, which would seem to represent more of an instrumental–pragmatic motivational profile than an integrative one in this new sense of identity as an “imagined identity” as a global citizen. Dörnyei (2006, pp. 53–54) suggests dispensing with the notion of integrative motivation altogether in favor of notions of identification and self-concept as regards a person's ideal self, in terms of hopes and dreams, and the person's “ought-to” self, in terms of satisfying external constraints and avoiding negative outcomes. In Dörnyei's (2006) self-based conception, “L2 motivation can be seen as the desire to reduce the perceived discrepancies between the learner's actual self and his/her ideal or ought-to L2 selves” (p. 54).⁶

Willingness to Communicate

An individual characteristic combining personality and other personal attributes with motivation and affective factors is **willingness to communicate** (WTC), applied to language learning by MacIntyre, Clement, Dörnyei, and Noels (1998). The WTC construct attempts to interrelate 12 linguistic, psychological, and social factors organized in six successive layers of influence on L2 performance. Dörnyei (2001) observes that “the model attempts to draw together a host of learner variables that have been well established as influences on second language acquisition and communication” (p. 254), including personality and self-confidence; communicative competence and experience; interpersonal motivation, desire to affiliate, and intergroup attitudes; and the social situation.

In such a complex model, the connection to pronunciation is not straightforward. A number of the WTC variables seem to be related to a number of other personal characteristics reviewed above that promote L2 communication, such as language learning aptitude, confidence, and proficiency; empathy, extraversion, and a multicultural personality; and low FLCA in general and specifically regarding pronunciation (i.e., low PA).⁷ WTC may then affect pronunciation as a consequence of learner agency to increase interaction and experience using the L2. Research by Derwing, Munro, and Thomson (2008) offers possible confirmation of this positive causal linkage for Slavic but not Chinese immigrants in Canada while also raising the possibility that WTC may vary according to cultural group. Although learner agency is clearly an important variable in language learning performance, further research is needed to clarify the extent to which pronunciation performance and achievement or ultimate attainment are predicted by such factors as willingness to communicate and to actively involve oneself in communication.

Learning Strategies

Some of the advantage of adults over children in terms of explicit learning can be explained by their having more, better, or more practiced learning strategies. As Taatgen (1999) notes, “learning strategies themselves have

to be learned” (p. 23); they can be developed over time as skilled behavior or “can themselves...be considered skills” (p. 212) which people access when trying to accomplish a goal. As opposed to cognitive or learning style, they are presumed to be more under the speaker’s control and thus a matter of agency. Individual differences in learning processes and outcomes are then in part a matter of the differential knowledge and application of learning strategies, which vary with the learner’s type of motivation (Oxford & Nykos, 1989; Richards, 1996; Schmidt et al., 1996; Schmidt & Watanabe, 2001). Specific learning strategies related to good pronunciation that have been researched and trialed (see Couper, 2011, 2015; Eckstein, 2007; Fraser, 2009; Szyszka, 2015) are reviewed in Chap. 4.

The Research Frontier: Brain Imaging

New research made possible by brain imaging has revealed differences in cognitive processing that are related to individual differences in both L1 and L2 language ability and achievement, including pronunciation, as well as factors of personality such as empathy that are represented by specific patterns of brain activation. Techniques such as **functional magnetic resonance imaging** (fMRI) are making it possible to study individual differences in pronunciation-related abilities—“pronunciation/speech imitation talent” (Reiterer et al., 2011)—directly at a cognitive level. Brain imaging research has shown that high-ability L2 learners have more efficient cognitive processing in frontal speech regions in perceiving and producing difficult L2 phonetic contrasts or novel speech sounds than low-ability learners (e.g., Golestani & Zatorre, 2004; Moser et al., 2009).

Reiterer et al.’s (2011) research has revealed individual differences in how late-onset adult L1 German learners of English, all of whom had learned English at around age 10, processed their L1, their L2, and additional brand new languages (Hindi and Tamil), labeled L0, in their left-hemisphere speech areas during imitation tasks. Higher ability imitators, according to native speaker assessments of their speech, had “enhanced gray matter volume” along with less activation “in a distinct fronto-parietal network,” where low-ability imitators showed comparatively higher

activation. The higher activation presumably reflects the greater difficulty of the task for the low-ability imitators, for whom the strength of their activation was lowest for L1, followed by L2, and then greatest for L0.

Two areas of the brain were found to be most relevant:

a premotor cluster, reflecting the speech motor execution of the articulatory movements (the “parroting part”) and second, the phonological loop mechanism of the acoustic working memory which integrates the phonological stream with the articulation output, located in the left inferior parietal area (the “phonology part”). The phonological loop is used for short term retention of verbal information and is a necessary prerequisite for later imitation of verbal material.... (Reiterer et al., 2011)

Reiterer et al. (2011) speculate that the involvement of these two clusters of brain activation in imitation and speech perception and production may involve the mirror neuron system that guides all kinds of imitative actions (Rizzolatti, Fogassi, & Gallese, 2001) and also seems to be the basis for theory of mind, which is crucial for language and other kinds of interactive behavior as well as empathy. Reiterer et al. (2011) note that an auditory mirror neuron system in the left hemisphere has been found to operate during auditorially triggered speech imitation in terms of exactly the two brain areas which were more active in their poor speech imitators.

Other research has shown involvement in the premotor regions for perception and production of prosody that correlates with empathy (Aziz-Zadeh, Sheng, & Gheyntanchi, 2010). Hu et al.’s (2013) research reinforces the findings of other studies suggesting key involvement of the mirror neuron system in speech imitation, which they also tested in their study of pronunciation aptitude, finding that the whole speech-motor network is involved in advanced learners’ L2 pronunciation, “including regions for speech-motor preparation/planning as well as speech-motor execution, and parts of the auditory-perceptual network, including the areas for perception of familiar/intelligible phonemes” (p. 374).

Reiterer et al. (2011) found a gender difference in that their German male participants were better imitators than their female participants, contrary to “the literature [which] attributed an advantage for second

language learning to females.” They speculate that besides a potential sampling bias or gender bias in German education or society, recent research suggests the possibility “that males have a significant advantage over females in motor skill learning.... Additionally, there is reported anecdotal superiority of males over females when rare and exceptional high talent in foreign language learning (including native-like accent) is concerned” (Reiterer et al., 2011), as an instance of the general statistical tendency for males to predominate at both the low and high extremes of all types of ability curves. Further research is obviously needed to explore gender and superior language skills and, specifically, imitative ability.

Concluding Remarks

The study of pronunciation in L2 learning is at an exciting stage of cross-fertilization across a number of fields of theoretical and applied study of language, psychology, and education, including cognitive linguistics, cognitive psychology, psycholinguistics, and the study of motivation. Key topics for continuing attention in the future are differences between L1 and L2 learning, individual differences, memory, and the functioning of the brain during learning and the performance of cognitive tasks having to do with pronunciation performance. Other topics ripe for further study are gender and culture in relation to individual differences and pronunciation learning.

Research on pronunciation learning and performance has demonstrated the centrality of both implicit and explicit learning and control of speech as well as the key roles played by L1 knowledge, identity, and individual motivations and characteristics such as aptitude and personality in accounting for the observed patterns of pronunciation behavior and achievement. These findings provide a valuable foundation of knowledge that gives direction for instruction and assessment of pronunciation and for pronunciation research in classrooms and other contexts, as explored in the chapters to come. Some notable practical initiatives related to language learning research that are reviewed in the next several chapters (Chaps. 3, 4, 5, and 6) are comparisons of methodologies that implement focus-on-form instruction; negotiation of input and feedback

on pronunciation in contexts of communication, perceptual training, and testing; instruction aimed at intelligibility and maintaining L1 culture and identity rather than aiming for accent-free pronunciation; application of learning strategies to enhance motivation and cater for individual differences; and use of electronic technologies for pronunciation assessment, input enhancement, and individualized instruction. In addition, the results of research on pronunciation learning and performance adds to a growing body of literature on the nature of pronunciation and its impacts on L2 communication that gives direction for continuing to investigate the effects of pronunciation on communication in different contexts, as explored in Chaps. 7 and 8.

In these different ways, pronunciation research and practice can be seen to be connected; and yet, as we emphasize in the final chapter (Chap. 8), there is still much in pronunciation practice—in language teaching, language testing, and language training in workplace contexts—that is insufficiently informed by and connected to research. In addition to reviewing in this book a wide range of research initiatives focused on applied work in pronunciation, we underscore the need to continue developing research and practice in interconnected ways, so that pronunciation teaching, testing, and training both informs and is supported by research.

Notes

1. Note that we do not in general make a distinction in meaning between the terms *acquisition* and *learning*. With regard to the traditional distinction made in SLA, it is our view that both acquisition (learning by means of implicit cognitive processing) and learning (learning by means of explicit and deliberate actions) are applicable to L1 and L2.
2. The limiting case is those who are severely deprived of linguistic input and practice throughout childhood, such as the neglected and abused child “Genie” (Curtiss, 1977) and the deaf child “E.M.” (Grimshaw, Adelstein, Bryden, & MacKinnon, 1998) and other deaf children with no sign language input or practice. The cause of failure to learn language in these cases may be lack of input alone or lack of input together with lack of practice producing language output. Children with certain cognitive, psy-

chological, or specifically phonological disorders (Chap. 7) are generally less severe cases but may also not learn to speak their mother tongue perfectly, or may be developmentally delayed in learning to speak it.

3. It can be noted that some of infants' spontaneous utterances (e.g., in nonsense babbling or involuntary emotive cries) may have no basis in their prior perceptual experience. It can be speculated that only in the case of those aspects of language which are instinctual or random are acts of production not based on prior related acts of perception.
4. Length of residence, in contrast, is less likely to be a proxy measure for amount of interaction or immersion in L2 environments since those who arrive after school age may be less likely to become immersed or integrated into an L2 environment, instead remaining relatively "sheltered" from L2 contexts in their own L1 circle or community (see below).
5. We note that this more explicit way of producing speech is a non-affiliative speaking style that expresses social distance and non-solidarity, and thus can affect the quality and quantity of input that those learners would receive from listeners.
6. Pennington (2018, p. 92) similarly speaks of reducing the gap between a language learner's aspirational identity and performed identity.
7. It can be observed that WTC is in some ways the converse or inverse of the unwillingness to talk or fear of speaking seen in FLCA or PA, so it should come as no surprise that WTC and PA were found to have a fairly strong negative correlation of $R = -0.60$ (Baran-Lucarz, 2014).

References

- Abrahamsson, N. (2012). Age of onset and nativelike L2 ultimate attainment of morphosyntactic and phonetic intuition. *Studies in Second Language Acquisition*, 34(2), 187–214. <https://doi.org/10.1017/S0272263112000022>
- Adjémian, C. (1976). On the nature of interlanguage systems. *Language Learning*, 26(2), 297–320. <https://doi.org/10.1111/j.1467-1770.1976.tb00279.x>
- Altenberg, E. P., & Vago, R. M. (1987/1983). Theoretical implications of an error analysis of second language phonology production. In G. Ioup & S. Weinberger (Eds.), *Interlanguage phonology: The acquisition of a second language sound system* (pp. 148–164). New York: Harper & Row. Originally published 1983 in *Language Learning*, 33(4), 427–448.

- Aziz-Zadeh, L., Sheng, T., & Gheyntanchi, A. (2010). Common premotor regions for the perception and production of prosody and correlations with empathy and prosodic ability. *PLoS One*, 5(1): e8759. <https://doi.org/10.1371/journal.pone.0008759>
- Baddeley, A. D. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, 63, 1–29. <https://doi.org/10.1146/annurev-psych-120710-100422>
- Ball, M. J. (2003). Clinical applications of a cognitive phonology. *Logopedics Phoniatrics Vocology*, 28(2), 63–69. <https://doi.org/10.1080/14015430310011763>
- Baran-Lucarz, M. (2012a). Individual learner differences and accuracy in foreign language pronunciation. In M. Pawlak (Ed.), *New perspectives on individual differences in language learning and teaching*. Berlin and Heidelberg: Springer.
- Baran-Lucarz, M. (2012b). Ego boundaries and attainments in FL pronunciation. *Studies in Second Language Learning and Teaching*, 2(1), 45–66. <https://doi.org/10.14746/ssl.2012.2.1.3>
- Baran-Lucarz, M. (2014). The link between pronunciation anxiety and willingness to communicate in the foreign-language classroom: The Polish EFL context. *Canadian Modern Language Review*, 70(4), 445–473. <https://doi.org/10.3138/cmlr.2666>
- Beebe, L. (1987/1980). Sociolinguistic variation and style shifting in second language acquisition. In G. Ioup & S. Weinberger (Eds.), *Interlanguage phonology: The acquisition of a second language sound system* (pp. 378–388). New York: Harper & Row. Originally published 1980 in *Language Learning*, 30(4), 433–447.
- Beebe, L. (1987/1984). Myths about interlanguage phonology. In G. Ioup & S. Weinberger (Eds.), *Interlanguage phonology: The acquisition of a second language sound system* (pp. 165–175). New York: Harper & Row. Originally published 1984 in S. Eliasson (Ed.), *Theoretical issues in contrastive phonology* (pp. 51–62). Heidelberg: Julius Groos Verlag.
- Best, C. T., & Tyler, M. D. (2007). Nonnative and second- language speech perception: Commonalities and complementaries. In M. J. Munro & O.-S. Bohn (Eds.), *Second language speech learning: The role of language experience in perception and production* (pp. 13–34). Amsterdam: John Benjamins.
- Bialystok, E. (1997). The structure of age: In search of barriers to second language acquisition. *Second Language Research*, 13(2), 116–137. <https://doi.org/10.1191/026765897677670241>
- Bialystok, E. (2002). On the reliability of robustness: A reply to DeKeyser. *Studies in Second Language Acquisition*, 24(3), 481–488. <https://doi.org/10.1017/S0272263102003054>

- Bialystok, E., Craik, F. I., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: Evidence from the Simon task. *Psychology and Aging, 19*(2), 290–303. <https://doi.org/10.1037/0882-7974.19.2.290>
- Bialystok, E., & Miller, B. (1999). The problem of age in second-language acquisition: Influences from language, structure, and task. *Bilingualism: Language and Cognition, 2*(2), 127–145. <https://doi.org/10.1017/S1366728999000231>
- Birdsong, D. (2005). Interpreting age effects in second language acquisition. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 109–127). New York: Oxford University Press.
- Birdsong, D. (2006). Age and second language acquisition and processing: A selective overview. *Language Learning, 56*(1), 9–49. <https://doi.org/10.1111/j.1467-9922.2006.00353.x>
- Birdsong, D. (2014). The critical period hypothesis for second language acquisition: Tailoring the coat of many colors. In M. Pawlak & L. Aronin (Eds.), *Essential topics in applied linguistics and multilingualism, second language learning and teaching* (pp. 43–50). Basel: Springer. https://doi.org/10.1007/978-3-319-01414-2_3
- Bley-Vroman, R. (1988). The fundamental character of foreign language learning. In W. Rutherford & M. Sharwood Smith (Eds.), *Grammar and second language teaching: A book of readings* (pp. 19–30). Rowley, MA: Newbury House.
- Bongaerts, T. (1999). Ultimate attainment in L2 pronunciation: The case of very advanced L2 learners. In D. Birdsong (Ed.), *Second language acquisition and the critical period hypothesis* (pp. 133–159). Mahwah, NJ: Lawrence Erlbaum.
- Bradford, B. (1997). Upspeak in British English. *English Today, 13*(3), 29–36. <https://doi.org/10.1017/S0266078400009810>
- Bradlow, A. (2008). Training non-native sound patterns: Lessons from training Japanese adults on the English /ɹ/ –/l/ contrast. In J. G. Hansen Edwards & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 287–308). Amsterdam and Philadelphia: John Benjamins.
- Brière, E. (1966). An investigation of phonological interference. *Language, 42*(4), 768–796. <https://doi.org/10.2307/411832>
- Brière, E. (1968). *A psycholinguistic study of phonological interference*. The Hague: Mouton.
- Browman, C., & Goldstein, L. (1995). Dynamics and articulatory phonology. In R. Port & T. van Gelder (Eds.), *Mind as motion: Explorations in the dynamics of cognition* (pp. 175–193). Cambridge, MA: MIT Press.

- Brown, A. (2008). Pronunciation and good language learners. In C. Griffiths (Ed.), *Lessons from good language learners* (pp. 197–207). Cambridge: Cambridge University Press.
- Brown, H. D. (1994). *Principles of language learning and teaching* (3rd ed.). New York: Prentice-Hall.
- Bybee, J. (2001). *Phonology and language use*. Cambridge: Cambridge University Press.
- Bybee, J. (2010). *Language, usage, and cognition*. Cambridge: Cambridge University Press.
- Carroll, J. B. (1981). Twenty-five years of research in foreign language aptitude: Then and now. In K. C. Diller (Ed.), *Individual differences and universals in language learning aptitude* (pp. 83–118). Rowley, MA: Newbury House.
- Carroll, J. B., & Sapon, S. (2002/1959). *Modern Language Aptitude Test*. San Antonio, TX and Bethesda, MD: Psychological Corporation and Second Language Teaching, Inc.
- Chapelle, C., & Roberts, C. (1986). Ambiguity tolerance and field independence as predictors of proficiency in English as a second language. *Language Learning*, 36(1), 27–45. <https://doi.org/10.1111/j.1467-1770.1986.tb00367.x>
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Clément, R., & Noels, K. A. (1992). Toward a situated approach to ethnolinguistic identity: The effects of status on individuals and groups. *Journal of Language and Social Psychology*, 11(4), 203–232. <https://doi.org/10.1177/0261927X92114002>
- 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. <https://doi.org/10.1080/09658416.2011.570347>
- Couper, G. (2015). Applying theories of language and learning to teaching pronunciation. In M. Reed & J. M. Levis (Eds.), *Handbook of English pronunciation* (pp. 413–432). New York: John Wiley & Sons.
- Croft, W., & Cruse, D. A. (2004). *Cognitive linguistics*. New York and Cambridge, UK: Cambridge University Press.
- Cruttenden, A. (1997/1986). *Intonation*. Cambridge, UK: Cambridge University Press.
- Curtiss, S. (1977). *Genie: A linguistic study of a modern day “wild child”*. New York: Academic Press.
- Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum Press.

- DeKeyser, R., Alfi-Shabta, I., & Ravid, D. (2010). Cross-linguistic evidence for the nature of age effects in second language acquisition. *Applied Psycholinguistics*, 31(3), 413–438. <https://doi.org/10.1017/S0142716410000056>
- DeKeyser, R., & Larson-Hall, J. (2005). What does the critical period really mean? In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 88–108). New York: Oxford University Press.
- Derwing, T. M., & Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching*, 42(4), 476–490. <https://doi.org/10.1017/S026144480800551X>
- Derwing, T. M., & Munro, M. J. (2013). The development of L2 oral language skills in two L1 groups: A seven-year study. *Language Learning*, 63(2), 163–185. <https://doi.org/10.1111/lang.12000>
- Derwing, T. M., Munro, M. J., & Thomson, R. I. (2008). A longitudinal study of ESL learners' fluency and comprehensibility development. *Applied Linguistics*, 29(3), 359–380. <https://doi.org/10.1093/applin/amm041>
- Derwing, T. M., Rossiter, M. J., Munro, M. J., & Thomson, R. I. (2004). L2 fluency: Judgments on different tasks. *Language Learning*, 54(4), 655–679. <https://doi.org/10.1111/j.1467-9922.2004.00282.x>
- Dewaele, J.-M. (2013). The link between foreign language classroom anxiety and psychoticism, extraversion, and neuroticism among adult bi- and multilinguals. *Modern Language Journal*, 97(3), 670–684. <https://doi.org/10.1111/j.1540-4781.2013.12036.x>
- Dewaele, J.-M., & Al-Saraj, T. M. (2015). Foreign Language Classroom Anxiety of Arab learners of English: The effect of personality, linguistic and sociobiographical variables. *Studies in Second Language Learning and Teaching*, 5(2), 205–228. <https://doi.org/10.14746/ssllt.2015.5.2.2>
- Dewaele, J.-M., & Furnham, A. (2000). Personality and speech production: A pilot study of second language learners. *Personality and Individual Differences*, 28(2), 355–365. [https://doi.org/10.1016/S0191-8869\(99\)00106-3](https://doi.org/10.1016/S0191-8869(99)00106-3)
- Dewaele, J.-M., & Li Wei. (2013). Is multilingualism linked to a higher tolerance of ambiguity? *Bilingualism: Language and Cognition*, 16(1), 231–240. <https://doi.org/10.1017/S1366728912000570>
- Dickerson, L. J. (1975). The learners' interlanguage as a system of variable rules. *TESOL Quarterly*, 9(4), 401–408. <https://doi.org/10.2307/3585624>
- Dickerson, W. (1977). Language variation in applied linguistics. *International Review of Applied Linguistics*, 35(1), 43–66. <https://doi.org/10.1075/itl.35.03dic>

- Djikic, M., Oatley, K., & Carland, M. (2012). Genre or artistic merit? The effect of literature on personality. *Scientific Study of Literature*, 2(1), 25–36. <https://doi.org/10.1075/ssol.2.1.02dji>
- Dörnyei, Z. (2001). *Teaching and researching motivation*. Harlow, UK and New York: Pearson Education.
- Dörnyei, Z. (2006). Individual differences in second language acquisition. *AILA Review*, 19, 42–68. <https://doi.org/10.1075/aila.19.05dor>
- Eckman, F. (1987/1977). Markedness and the contrastive analysis hypothesis. In G. Ioup & S. Weinberger (Eds.), *Interlanguage phonology: The acquisition of a second language sound system* (pp. 55–69). New York: Harper & Row. Originally published 1977 in *Language Learning*, 27(2), 315–330.
- Eckman, F. (1987/1981). On the naturalness of interlanguage phonological rules. In G. Ioup & S. Weinberger (Eds.), *Interlanguage phonology: The acquisition of a second language sound system* (pp. 125–144). New York: Harper & Row. Originally published 1981 in *Language Learning*, 31(1), 195–216.
- Eckstein, G. T. (2007). *A correlation of pronunciation learning strategies with spontaneous English pronunciation of adult ESL learners*. MA thesis, Brigham Young University. Cited from Szyszka (2015).
- Ehrman, M. E., & Oxford, R. L. (1995). Cognition plus: Correlates of language learning success. *Modern Language Journal*, 79(1), 67–89. <https://doi.org/10.2307/329394>
- Ellis, N. C. (1994). Implicit and explicit processes in language acquisition: An introduction. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 1–32). London: Academic Press.
- Ellis, N. C. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition*, 18(1), 91–126. <https://doi.org/10.1017/S0272263100014698>
- Ellis, N. C. (2001). Memory for language. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 33–68). Cambridge: Cambridge University Press.
- Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24(2), 143–188. <https://doi.org/10.1017/S0272263102002024>
- Ellis, N. C. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition*, 27(2), 305–352. <https://doi.org/10.1017/S027226310505014X>
- Escudero, P. (2005). *Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization*. PhD disserta-

- tion, University of Utrecht. Utrecht: LOT (Netherlands Graduate School of Linguistics).
- Escudero, P. (2007). Second-language phonology: The role of perception. In M. C. Pennington (Ed.), *Phonology in context* (pp. 109–134). Basingstoke, Hampshire, UK, and New York: Palgrave Macmillan.
- Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*, 26(4), 551–585. <https://doi.org/10.1017/S0272263104040021>
- Fernald, A., Truet, T., Dunn, J., Papansek, M., Boysson, B., & Fuleni, I. (1989). A cross language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*, 16(3), 477–501. <https://doi.org/10.1017/S0305000900010679>
- Flege, J. E. (1987). The production of “new” and “similar” phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics*, 15(1), 47–65. Retrieved January 1, 2018, from <https://pdfs.semanticscholar.org/4a01/a0def4d4016c6fec441b3f2005087e0041e2.pdf>
- Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–276). Timonium, MD: York Press.
- Flege, J. E. (2003). Assessing constraints on second-language segmental production and perception. In A. Meyer & N. Schiller (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 319–355). Berlin: Mouton de Gruyter.
- Flege, J. E. (2007). Language contact in bilingualism: Phonetic system interactions. In J. Cole & J. Hualde (Eds.), *Laboratory Phonology 9* (pp. 353–380). Berlin: Mouton de Gruyter.
- Flege, J. E. (2009). Give input a chance! In T. Piske & M. Young-Scholten (Eds.), *Input matters in SLA* (pp. 175–190). Bristol, Buffalo and Toronto: Multilingual Matters.
- Flege, J. E., & Eefting, W. (1986). Linguistic and developmental effects on the production and perception of stop consonants. *Phonetica*, 43(4), 155–171. <https://doi.org/10.1159/000261768>
- Flege, J. E., & Eefting, W. (1987). Cross-language switching in stop consonant production and perception by Dutch speakers of English. *Speech Communication*, 6(3), 185–202. [https://doi.org/10.1016/0167-6393\(87\)90025-2](https://doi.org/10.1016/0167-6393(87)90025-2)
- Flege, J. E., Frieda, E. M., & Nozawa, T. (1997). Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics*, 25(2), 169–186. <https://doi.org/10.1006/jpho.1996.0040>

- Flege, J. E., & Hillenbrand, J. (1987/1984). Limits on pronunciation accuracy in adult foreign language speech production. In G. Ioup & S. Weinberger (Eds.), *Interlanguage phonology: The acquisition of a second language sound system* (pp. 176–203). New York: Harper & Row. Originally published in *Journal of the Acoustical Society of America*, 76(3), 708–721.
- Flege, J. E., Munro, M., & MacKay, I. R. A. (1995). Factors affecting degree of perceived foreign accent in a second language. *Journal of the Acoustical Society of America*, 97(5), 3125–3134. <https://doi.org/10.1121/1.413041>
- Flege, J. E., Yeni-Komshian, G., & Liu, S. (1999). Age constraints on second language acquisition. *Journal of Memory & Language*, 41(1), 78–104. <https://doi.org/10.1006/jmla.1999.2638>
- Fought, C. (2002). Ethnicity. In J. K. Chambers, P. Trudgill, & N. Schilling-Estes (Eds.), *The handbook of language variation and change* (pp. 444–472). Oxford: Blackwell.
- Fought, C. (2006). *Language and ethnicity: Key topics in sociolinguistics*. New York: Cambridge University Press.
- Fraser, H. (2009). Pronunciation as categorization: The role of contrast in teaching English /t/ and /l/. In A. Mahboob & C. Lipovsky (Eds.), *Studies in applied linguistics and language learning* (pp. 289–306). Newcastle upon Tyne: Cambridge Scholars Publishing.
- Gardner, R. C. (1982). Language attitudes and language learning. In E. Bouchard Ryan & H. Giles (Eds.), *Attitudes towards language variation* (pp. 132–147). London: Edward Arnold.
- Gardner, R. C. (1985). *Social psychology and second language learning: The role of attitudes and motivation*. London: Edward Arnold.
- Gardner, R. C. (2007, June). Motivation and second language acquisition. *Porta Linguarum*, 8, 9–20. Retrieved on January 18, 2018, from digibug.ugr.es/bitstream/10481/31616/1/Gardner.pdf. Also available in an earlier (Dec 15, 2006) version at publish.uwo.ca/~gardner/docs/SPAIN_TALK.pdf
- Gatbonton, E., & Trofimovich, P. (2008). The ethnic group affiliation and L2 proficiency link: Empirical evidence. *Language Awareness*, 17(3), 229–248. <https://doi.org/10.1080/09658410802146867>
- Gatbonton, E., Trofimovich, P., & Segalowitz, N. (2011). Ethnic group affiliation and patterns of development of a phonological variable. *Modern Language Journal*, 95(2), 188–204. <https://doi.org/10.1111/j.1540-4781.2011.01177.x>
- Giles, H. (1979). Ethnicity markers in speech. In K. R. Scherer & H. Giles (Eds.), *Social markers in speech* (pp. 251–290). Cambridge: Cambridge University Press.

- Giles, H., & Byrne, J. L. (1982). An intergroup approach to second language acquisition. *Journal of Multilingual and Multicultural Development*, 3(1), 17–40. <https://doi.org/10.1080/01434632.1982.9994069>
- Golestani, N., & Zatorre, R. J. (2004). Learning new sounds of speech: Reallocation of neural substrates. *Neuroimage*, 21(2), 494–506. <https://doi.org/10.1016/j.neuroimage.2003.09.071>
- Granena, G., & Long, M. H. (2013a). Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research*, 29(3), 311–343. <https://doi.org/10.1177/0267658312461497>
- Granena, G., & Long, M. H. (Eds.). (2013b). *Sensitive periods, language aptitudes, and ultimate L2 attainment*. Amsterdam: John Benjamins.
- Green, K. P., Zampini, M. L., & Magloire, J. (1997). An examination of word-initial stop closure interval in English, Spanish and Spanish-English bilinguals. *Journal of the Acoustical Society of America*, 102(5), 3136. <https://doi.org/10.1121/1.420648>
- Grimshaw, G. M., Adelstein, A., Bryden, M. P., & MacKinnon, G. E. (1998). First-language acquisition in adolescence: Evidence for a critical period for verbal language development. *Brain and Language*, 63(2), 237–255. <https://doi.org/10.1006/brln.1997.1943>
- Guiora, A. (1972). Construct validity and transpositional research: Toward an empirical study of psychoanalytic concepts. *Comprehensive Psychiatry*, 13(2), 139–150. [https://doi.org/10.1016/0010-440X\(72\)90019-3](https://doi.org/10.1016/0010-440X(72)90019-3)
- Guiora, A., Acton, W., Erard, R., & Strickland, F. (1980). The effects of benzodiazepine (Valium) on permeability of language ego boundaries. *Language Learning*, 30(2), 351–363. <https://doi.org/10.1111/j.1467-1770.1980.tb00323.x>
- Guiora, A., Beit-Hallahmi, H., Brannon, R., Dull, C., & Scovel, T. (1972). The effects of experimentally induced changes in ego states on pronunciation ability in a second language: An exploratory study. *Comprehensive Psychiatry*, 13(5), 421–428. [https://doi.org/10.1016/0010-440X\(72\)90083-1](https://doi.org/10.1016/0010-440X(72)90083-1)
- Guiora, A., Lane, H. L., & Bosworth, L. A. (1967). An exploration of some personality variables in authentic pronunciation of a second language. In H. L. Lane & E. M. Zale (Eds.), *Studies in language and language behavior progress report IV*. Ann Arbor, MI: Center for Research on Language and Language Behavior.
- Gumperz, J. (Ed.). (1982). *Language and social identity*. New York: Cambridge University Press.

- Hakuta, K., Bialystok, E., & Wiley, E. (2003). Critical evidence: A test of the critical-period hypothesis for second-language acquisition. *Psychological Science*, 14(1), 31–38. <https://doi.org/10.1111/1467-9280.01415>
- Hansen, J., & Stansfield, C. (1981). The relationship of field dependent-independent cognitive styles to foreign language achievements. *Language Learning*, 31(2), 349–367. <https://doi.org/10.1111/j.1467-1770.1981.tb01389.x>
- Hansen, L. (1984). Field dependence-independence and language testing: Evidence from six Pacific island cultures. *TESOL Quarterly*, 18(2), 311–324. <https://doi.org/10.2307/3586696>
- Hansen Edwards, J. G. (2008). Social factors and variation in production of L2 phonology. In J. G. Hansen Edwards & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 251–279). Amsterdam and Philadelphia: John Benjamins.
- Hazan, V. L., & Boulaïka, G. (1993). Perception and production of a voicing contrast by French-English bilinguals. *Language and Speech*, 36(1), 17–38. <https://doi.org/10.1177/002383099303600102>
- Hecht, B. F., & Mulford, R. (1987/1980). The acquisition of a second language phonology: Interaction of transfer and developmental factors. In G. Ioup & S. Weinberger (Eds.), *Interlanguage phonology: The acquisition of a second language sound system* (pp. 213–228). New York: Harper & Row. Originally published 1980 in *Papers and Reports in Child Language Development*, 18, 61–74.
- Holzman, M. (1997). *The language of children* (2nd ed.). Cambridge, MA and Oxford, UK: Blackwell.
- Hopp, H., & Schmid, M. (2013). Perceived foreign accent in first language attrition and second language acquisition: The impact of age of acquisition and bilingualism. *Applied Psycholinguistics*, 34(2), 361–394. <https://doi.org/10.1017/S0142716411000737>
- Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal*, 70(2), 125–132. <https://doi.org/10.2307/327317>
- Hu, X., Ackermann, J., Martin, J. A., Erb, M., Winkler, S., & Reiterer, S. M. (2013). Language aptitude for pronunciation in advanced second language (L2) Learners: Behavioural predictors and neural substrates. *Brain and Language*, 127(3), 366–376. <https://doi.org/10.1016/j.bandl.2012.11.006>
- Hutnik, N. (1991). *Ethnic minority identity: A social psychological perspective*. Oxford: Oxford University Press.

- Iverson, P., & Kuhl, P. K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of the Acoustical Society of America*, 97(1), 553–562. <https://doi.org/10.1121/1.412280>
- Jensen, F. E. (2015, September/October). Plastic fantastic. *Pennsylvania Gazette*, 56–62.
- Johansson, F. A. (1973). *Immigrant Swedish phonology: A study in multiple contact analysis*. Lund: CWK Gleerup. Cited in Beebe (1987/1984).
- Johnson, J., & Newport, E. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of ESL. *Cognitive Psychology*, 21(1), 60–99. Retrieved January 1, 2018, from <https://msu.edu/~ema/803/Ch12-LanguageStructure/1/JohnsonNewport89.pdf>. Also available at <https://www.sciencedirect.com/journal/cognitive-psychology/vol/21/issue/1>
- Kerswill, P., & Shockey, L. (2007). The description and acquisition of variable phonological patterns: Phonology and sociolinguistics. In M. C. Pennington (Ed.), *Phonology in context* (pp. 51–75). Basingstoke, Hampshire, UK and New York: Palgrave Macmillan.
- Kuhl, P. K. (2000). A new view of language. *Colloquium. PNAS: Publications of the National Academy of Sciences*, 97(22), 11850–11857. <https://doi.org/10.1073/pnas.97.22.11850>
- Kuhl, P., & Iverson, P. (1995). Linguistic experience and the “perceptual magnet effect”. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross language research* (pp. 121–154). Timonium, MD: York Press.
- Labov, W. (2001). *Principles of linguistic change: Social factors*. Oxford, UK and Cambridge, MA: Blackwell.
- Lado, R. (1957). *Linguistics across cultures: Applied linguistics for language teachers*. Ann Arbor: University of Michigan Press.
- Larsen-Freeman, D., & Long, M. H. (1991). *An introduction to second language acquisition research*. London and New York: Longman.
- Leather, J. (1999). Second-language speech research: An introduction. *Language Learning*, 49(s1), 1–56. <https://doi.org/10.1111/0023-8333.49.s1.1>
- Lenneberg, E. H. (1967). *Biological foundations of language*. New York: Wiley.
- Long, M. H. (1981). Input, interaction and second language acquisition. In H. Winitz (Ed.), *Native and foreign language acquisition* (pp. 259–278). Annals of the New York Academy of Sciences, 379. New York: New York Academy of Sciences. <https://doi.org/10.1111/j.1749-6632.1981.tb42014.x>
- Long, M. H. (1983). Native speaker/non-native speaker conversation and the negotiation of comprehensible input. *Applied Linguistics*, 4(2), 126–141. <https://doi.org/10.1093/applin/4.2.126>

- Long, M. H. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition*, 12(3), 251–285. <https://doi.org/10.1017/S0272263100009165>
- Long, M. H. (1996). The role of linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). New York: Academic Press.
- Long, M. (2015). *Second language acquisition and task-based language teaching*. Malden, MA & Oxford, UK: Wiley Blackwell.
- MacIntyre, P. D., Clément, R., Dörnyei, Z., & Noels, K. A. (1998). Conceptualizing willingness to communicate in a L2: A situational model of L2 confidence and affiliation. *Modern Language Journal*, 82(4), 545–562. <https://doi.org/10.2307/330224>
- Major, R. C. (1987). A model for interlanguage phonology. In G. Ioup & S. H. Weinberger (Eds.), *Interlanguage phonology: The acquisition of a second language sound system* (pp. 101–124). Cambridge, MA: Newbury House.
- Major, R. C. (1992). Losing English as a first language. *The Modern Language Journal*, 76(2), 190–208. <https://doi.org/10.2307/329772>
- Major, R. C. (2001). *Foreign accent: The ontogeny and phylogeny of second language speech*. Mahwah, NJ: Lawrence Erlbaum.
- Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science*, 283(5398), 77–80. <https://doi.org/10.1126/science.283.5398.77>
- Maye, J., & Gerken, L. A. (2000). Learning phonemes without minimal pairs. In S. C. Howell, S. A. Fish, & T. Keith-Lucas (Eds.), *Proceedings of the 24th Annual Boston University conference on language development* (pp. 522–533). Somerville, MA: Cascadilla Press.
- Meara, P. (2005). *LLAMA language aptitude tests*. Swansea, UK: Lognostics, University of Wales Swansea. Retrieved January 1, 2018, from http://www.lognostics.co.uk/tools/llama/llama_manual.pdf
- Miettinen, H. (2012). *Phonological working memory and L2 knowledge: Finnish children learning English*. Jyväskylä Studies in Humanities, 184. Academic Dissertation, University of Jyväskylä. Retrieved from jyx.jyu.fi/dspace/bitstream/handle/123456789/38386/978-951-39-4825-2.pdf?sequence=1
- Mompean, J. A. (2014). Phonology. In J. R. Taylor & J. Littlemore (Eds.), *Bloomsbury companion to cognitive linguistics* (pp. 357–392). London: Bloomsbury.
- Moser, D., Fridriksson, J., Bonilha, L., Healy, E. W., Baylis, G., Baker, J. M., et al. (2009). Neural recruitment for the production of native and novel

- speech sounds. *Neuroimage*, 46(2), 549–557. <https://doi.org/10.1016/j.neuroimage.2009.01.015>
- Moyer, A. (2004). *Age, accent and experience in second language acquisition: An integrated approach to critical period inquiry*. Clevedon: Multilingual Matters.
- Naiman, N. M., Fröhlich, M., & Stern, H. H. (1978). *The good language learner*. Toronto: Ontario Institute for Studies in Education.
- Nemser, W. (1971). *An experimental study of phonological interference in the English of Hungarians*. Bloomington, IN: Indian University Press.
- Oller, D. K. (2000). *The emergence of the speech capacity*. Mahwah, NJ: Lawrence Erlbaum.
- Oxford, R., & Nykos, M. (1989). Variables affecting choice of language learning strategies by university students. *Modern Language Journal*, 73(3), 291–300. <https://doi.org/10.2307/327003>
- Oyama, S. (1976). A sensitive period for the acquisition of a non-native phonological system. *Journal of Psycholinguistic Research*, 5(3), 261–283. <https://doi.org/10.1007/BF01067377>
- Patkowski, M. (1990). Age and accent in a second language: A reply to James Emil Flege. *Applied Linguistics*, 11(1), 73–89. <https://doi.org/10.1093/applin/11.1.73>
- Pennington, M. C. (2002). Equivalence classification in L2 phonology: In search of the mechanisms. In J. Leather & A. James (Eds.), *New sounds 2000, Proceedings of the Fourth International Symposium on the Acquisition of Second-Language Speech*, University of Amsterdam, September 2000 (pp. 280–289). Klagenfurt, Austria: University of Klagenfurt.
- Pennington, M. C. (2018). Identity in language learning. In J. C. Richards & A. Burns (Eds.), *The Cambridge guide to learning English as a second language* (pp. 91–98). New York: Cambridge University Press.
- Pennington, M. C. (forthcoming). *Introduction to bilingualism and multilingualism: People and languages in contact*. Malden, MA and Chichester, UK: Wiley-Blackwell.
- Pennington, M. C., Lau, L., & Sachdev, I. (2011). Diversity in adoption of linguistic features of London English by Chinese and Bangladeshi adolescents. *Language Learning Journal*, 39(2), 177–199. <https://doi.org/10.1080/09571736.2011.573686>
- Pervin, L. A., & Cervone, D. (2010). *Personality: Theory and research*. Hoboken, NJ: John Wiley.
- Piske, T., & MacKay, I. R. A. (1999). Age and L1 use effects on degree of foreign accent in English. In *Proceedings of the 14th International Congress of*

- Phonetic Sciences* (pp. 1433–1436), August 1–7. San Francisco. Retrieved January 1, 2018, from http://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS1999/papers/p14_1433.pdf
- Purcell, E., & Suter, R. (1980). Predictors of pronunciation accuracy: A reexamination. *Language Learning*, 30(2), 271–287. <https://doi.org/10.1111/j.1467-1770.1980.tb00319.x>
- Rassaei, E. (2015). Recasts, field dependence/independence cognitive style, and L2 development. *Language Teaching Research*, 19(4), 499–518. <https://doi.org/10.1177/1362168814541713>
- Reis, A., & Castro-Caldas, A. (1997). Illiteracy: A cause for biased cognitive development. *Journal of the International Neuropsychological Society*, 3(5), 444–450. Retrieved August 15, 2017, from <https://www.cambridge.org/core/journals/journal-of-the-international-neuropsychological-society/article/illiteracy-a-cause-for-biased-cognitive-development/880830EAC2B2A5778A33C7F450>
- Reiterer, S. M., Hu, X., Erb, M., Rota, G., Nardo, D., Grodd, W., et al. (2011, October). Individual differences in audio-vocal speech imitation aptitude in late bilinguals: Functional neuro-imaging and brain morphology. *Frontiers in Psychology*, 2(article 271), 1–12. <https://doi.org/10.3389/fpsyg.2011.00271>
- Reiterer, S. M., Singh, N. C., & Winkler, S. (2012). Predicting speech imitation ability biometrically. In B. Stolterfoht & S. Featherstone (Eds.), *Current work in linguistic evidence: The fourth Tübingen meeting* (pp. 317–340). Berlin/New York: Mouton de Gruyter.
- Richards, J. C. (1974/1971). A non-contrastive approach to error analysis. In J. C. Richards (Ed.), *Error analysis: Perspectives on second language acquisition* (pp. 172–188). London: Longman. Originally published 1971 in *English Language Teaching*, 25(3), 204–219.
- Richards, S. (1996). Learning English in Hong Kong: Making connections between motivation, language use, and strategy choice. In M. C. Pennington (Ed.), *Language in Hong Kong at century's end* (pp. 303–338). Hong Kong: Hong Kong University Press.
- Rimfield, K., Dale, P. S., & Plomin, R. (2015). How specific is second language-learning ability? A twin study exploring the contributions of first language achievement and intelligence to second language achievement. *Translational Psychiatry*, 5(9), e638. <https://doi.org/10.1038/tp.2015.128>
- Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Neuroscience Reviews*, 2(9), 661–670. <https://doi.org/10.1038/35090060>

- Rochet, B. (1995). Perception and production of second-language speech sounds by adults. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language speech research* (pp. 379–410). Timonium, MD: York Press.
- Sachdev, I., & Bourhis, R. Y. (2005). Multilingual communication and social identification. In J. Harwood & H. Giles (Eds.), *Intergroup communication: Multiple perspectives* (pp. 65–92). New York: Peter Lang.
- Sachdev, I., & Giles, H. (2004). Bilingual accommodation. In T. K. Bhatia & W. C. Ritchie (Eds.), *Handbook of bilingualism* (pp. 353–378). Oxford: Blackwell.
- Sáfár, A. & Kormos, J. (2008). Revisiting problems with foreign language aptitude. *IRAL—International Review of Applied Linguistics in Language Teaching*, 46(2), 113–136. <https://doi.org/10.1515/IRAL.2008.005>. Retrieved January 1, 2018, from http://www.researchgate.net/publication/228848419_Revisiting_problems_with_foreign_language_aptitude
- Saffran, J. R., Aslin, R., & Newport, E. (1996). Statistical learning by 8-month old infants. *Science*, 274, 1926–1928. <https://doi.org/10.1126/science.274.5294.1926>
- Saito, K. (2015a). Experience effects on the development of late second language learners' oral proficiency. *Language Learning*, 65(3), 563–595. <https://doi.org/10.1111/lang.12120>
- Saito, K. (2015b). The role of age of acquisition in late second language oral proficiency attainment. *Studies in Second Language Acquisition*, 37(4), 713–743. <https://doi.org/10.1017/S0272263115000248>
- Saito, K. (2017). Effects of sound, vocabulary and grammar learning aptitude on adult second language speech attainment in foreign language classrooms. *Language Learning*, 67(3), 665–693. <https://doi.org/10.1111/lang.12244>.
- Sardegna, V. G., Lee, J., & Kusey, C. (2014). Development and validation of the learner attitudes and motivations for pronunciation (LAMP) inventory. *System*, 47(2), 162–175. <https://doi.org/10.1016/j.system.2014.10.009>
- Schmidt, R. W. (1983). Interaction, acculturation, and the acquisition of communicative competence: A case study of an adult. In N. Wolfson & E. Judd (Eds.), *Sociolinguistics and second language acquisition* (pp. 137–174). Rowley, MA: Newbury House.
- Schmidt, R. W. (1987/1977). Sociolinguistic variation and language transfer in phonology. In G. Ioup & S. Weinberger (Eds.), *Interlanguage phonology: The acquisition of a second language sound system* (pp. 365–377). New York: Harper

- & Row. Originally published 1977 in *Working Papers in Bilingualism*, 12, 79–95.
- Schmidt, R. W., Boraie, D., & Kassabgy, O. (1996). Foreign language motivation: International structure and external connections. In R. Oxford (Ed.), *Language learning motivation: Pathways to the new century*. Technical Report #11 (pp. 9–70). Honolulu: University of Hawai'i, Second Language Teaching & Curriculum Center. Retrieved June 1, 2017, from <http://nflrc.hawaii.edu/PDFs/SCHMIDT%20Foreign%20language%20motivation.pdf>
- Schmidt, R. W., & Watanabe, Y. (2001). Motivation, strategy use, and pedagogical preferences in foreign language learning. In Z. Dörnyei & R. Schmidt (Eds.), *Motivation and second language acquisition*. Technical Report #23 (pp. 313–359). Honolulu: University of Hawai'i, Second Language Teaching and Curriculum Center. Retrieved June 16, 2017, from <http://nflrc.hawaii.edu/PDFs/SCHMIDT%20Motivation,%20strategy%20use,%20and%20pedagogical%20preferences.pdf>
- Schumann, J. (1978). *The pidginization process: A model for second language acquisition*. Rowley, MA: Newbury House.
- Scovel, T. (2000). A critical review of the critical period research. *Annual Review of Applied Linguistics*, 20, 213–223. <https://doi.org/10.1017/S0267190500200135>
- Selinker, L. (1974/1972). Interlanguage. In J. C. Richards (Ed.), *Error analysis: Perspectives on second language acquisition* (pp. 31–54). London: Longman. Originally published 1972 in *IRAL: International Review of Applied Linguistics in Language Teaching*, 10(1–4), 209–232. <https://doi.org/10.1515/iral.1972.10.1-4.209>
- Sheldon, A., & Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics*, 3(3), 243–261. <https://doi.org/10.1017/S0142716400001417>
- Skehan, P. (1998). *A cognitive approach to language learning*. Oxford and New York: Oxford University Press.
- Smit, U. (2002). The interaction of motivation and achievement in advanced EFL pronunciation learners. *IRAL: International Review of Applied Linguistics*, 40(2), 89–116. <https://doi.org/10.1515/iral.2002.009>
- Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning*, 60(2), 263–308. <https://doi.org/10.1111/j.1467-9922.2010.00562.x>
- Stockwell, R., & Bowen, J. (1965). *The sounds of English and Spanish*. Chicago: University of Chicago Press.

- Strange, W., & Shafer, V. L. (2008). Speech perception in second language learners: The re-education of selective perception. In J. G. Hansen Edwards & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 153–191). Amsterdam and Philadelphia: John Benjamins.
- Szyszka, M. (2015). Good English pronunciation users and their pronunciation learning strategies. *Research in Language*, 13(1), 93–106. <https://doi.org/10.1515/rela-2015-0017>
- Taatgen, N. A. (1999). *Learning without limits: From problem solving towards a universal theory of learning*. PhD thesis, University of Groningen. Retrieved August 16, 2017, from [http://www.rug.nl/research/portal/publications/pub\(884cfcabd-a584-4e39-90c0-5452e13aeae\).html](http://www.rug.nl/research/portal/publications/pub(884cfcabd-a584-4e39-90c0-5452e13aeae).html)
- Tarone, E. (1979). Interlanguage as chameleon. *Language Learning*, 29(1), 181–191. <https://doi.org/10.1111/j.1467-1770.1979.tb01058.x>
- Tarone, E. (1987/1978). The phonology of interlanguage. In G. Ioup & S. Weinberger (Eds.), *Interlanguage phonology: The acquisition of a second language sound system* (pp. 70–85). New York: Harper & Row. Originally published 1978 in J. C. Richards (Ed.), *Understanding second and foreign language learning* (pp. 15–33). Rowley, MA: Newbury House.
- Tarone, E. (1988). *Variation in interlanguage*. London: Arnold.
- Tarone, E., Bigelow, M., & Hansen, K. (2009). *Literacy and second language oracy*. Oxford and New York: Oxford University Press.
- Thelen, E. (2014/1991). Motor aspects of emergent speech: A dynamic approach. In N. A. Krasnegor, D. M. Rumbaugh, R. L. Schiefelbusch, & M. Studdert-Kennedy (Eds.), *Biological and behavioral determinants of language development* (pp. 339–362). New York and London: Psychology Press. Originally published in 1991 by Hove & London: Lawrence Erlbaum Associates.
- Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA and London, UK: Harvard University Press.
- Van der Zee, K., van Oudenhoven, J. P., Ponterotto, J. G., & Fietzer, A. W. (2013). Multicultural personality questionnaire: Development of a short form. *Journal of Personality Assessment*, 95(1), 118–124. <https://doi.org/10.1080/00223891.2012.718302>
- Velleman, S. I., & Vihman, M. M. (2007). Phonology in infancy and childhood: Implications for theories of language learning. In M. C. Pennington (Ed.), *Phonology in context* (pp. 25–50). Basingstoke, Hampshire, UK and New York: Palgrave Macmillan.

- Weinberger, S. H. (1987). The influence of linguistic context on syllable simplification. In G. Ioup & S. Weinberger (Eds.), *Interlanguage phonology: The acquisition of a second language sound system* (pp. 401–417). New York: Harper & Row.
- Witkin, H. A., & Goodenough, D. R. (1981). *Cognitive styles: Essence and origins, field dependence and field independence*. New York: International Universities Press.
- Witkin, H. A., Oltman, P., Raskin, E., & Karp, S. (1971). *A manual for The Group Embedded Figures Test*. Palo Alto: Consulting Psychologists Press.
- Wode, H. (1977). The L2 acquisition of /r/. *Phonetica*, 34(3), 200–217. <https://doi.org/10.1159/000259877>
- Wode, H. (1983). Phonology in L2 acquisition. In H. Wode (Ed.), *Papers on language acquisition, language learning and language teaching* (pp. 175–187). Heidelberg: Julius Groos Verlag.
- Wulff, S., & Ellis, N. C. (2018). Usage-based approaches to SLA. In D. Miller, F. Bayram, J. Rothman, & L. Serratrice (Eds.), *Bilingual cognition and language: The state of the science across its subfields* (pp. 37–56). Amsterdam and Philadelphia: John Benjamins.
- Zampini, M. L. (2008). L2 speech production research: Findings, issues, and advances. In J. G. Hansen Edwards & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 219–249). Amsterdam and Philadelphia: John Benjamins.
- Zampini, M. L., & Green, K. P. (2001). The voicing contrast in English and Spanish: The relationship between production and perception. In J. Nicol (Ed.), *Two languages, one mind: Bilingual language processing* (pp. 23–48). Cambridge, MA: Blackwell.
- Zuengler, J. (1982). Applying accommodation theory to variable performance data in L2. *Studies in Second Language Acquisition*, 4(2), 181–192. <https://doi.org/10.1017/S0272263100004411>
- Zuengler, J. (1988). Identity markers and L2 pronunciation. *Studies in Second Language Acquisition*, 10(1), 33–50. <https://doi.org/10.1017/S027226310000694X>



3

Framing the Teaching of Pronunciation

Introduction

As with other areas of language, pronunciation teaching has historically been influenced by various trends in pedagogical approaches resulting in shifts in focus in teaching priorities and concerns. The natural evolution of spoken English, together with changing patterns in the use of English, particularly the increasingly dominant function of English as a language for international and intercultural communication, mean that it is useful to review the role and processes of pronunciation teaching in a historical and contemporary context, as a basis for decision-making to ensure that this area of the language curriculum offers good value in relation students' real-life needs.

In this chapter, we situate pronunciation teaching within the historical development of language theory and pedagogy and of the diverse and ever-widening contexts of English, and consider how such developments feed into practice to influence key decisions about appropriate models, goals, and priorities for pronunciation teaching and learning. The information covered here has direct practical value in offering a systematic basis for teachers to make curriculum decisions in this area of language pedagogy. To this end,

the chapter provides a set of pedagogical considerations, along with conceptual and factual background, that continue to build awareness and the relevant knowledge base linking pronunciation theory, research, and practice in the project begun in Chap. 1. It also offers two case studies of research illustrating ways in which learner needs and the context in which they are using English as L2 might feed into teaching practices as a form of applied research. In this way, instructional approaches are supported by research findings and empirical research is carried out in contexts of instruction, thereby strengthening research-practice, as well as theory-practice, connections.

The Changing Place of Pronunciation in Language Teaching

The place of pronunciation in language teaching has changed over time, impacted by changing conceptions of language and language learning as reviewed in Chap. 2. Adapting from the discussion in Pennington (2015a, 2015b), we view the changing position of pronunciation as a reflection of a historical shift of language teaching in the second half of the twentieth century from a formal language emphasis that gave pronunciation a central position, to a communicative emphasis that sidelined pronunciation within both second language and bilingual approaches to language teaching. Later, pronunciation was reinstated in second- language (L2) teaching as an aspect of a focus on form within a focus on meaning and communication, while bilingual and multilingual approaches to language teaching continued to downplay the importance of pronunciation for communication. A growing emphasis in contemporary multilingual and plurilingual orientations is on pronunciation as a reflection of speaker agency, identity, and affiliation.

1950s–1960s: Structural Language Teaching and Audiolingualism

In the era of the 1950s and 1960s, the dominant theory of language was structuralism (Harris, 1951), centered on formal grammar, and matched to the learning theory of behaviorism (Skinner, 1957), centered on habit

formation. Out of this combined emphasis in linguistics and psychology grew language teaching methodology focused on language structure and the language teaching approach of **Audiolingualism** that relied heavily on drilling and repetition of oral language (e.g., English Language Services, 1964; Lado, 1964) performed chorally in class and also individually in the language laboratory using audiotaped materials and headphones. Since pronunciation was considered a deeply ingrained and automatized feature of language competence that needed intensive instruction in order to alter long-established habits, a main focus of language teaching was repetition drills centered on individual phonemes and the prosody of grammatical phrases and sentences, with a goal of accurate perception and production of L2 pronunciation according to a standard-language, native speaker target. By the mid-1960s, structural linguistics and behaviorism were losing ground to Chomsky's (1965) notions of transformational grammar and linguistic innateness, and were essentially abandoned by the end of the 1960s.

1970s–1980s: Language Acquisition Processes and Communicative Language Teaching

In the 1970s and 1980s, the view of language broadened to incorporate language functions (Halliday, 1970, 1973) and social context (Labov, 1972), and not just grammatical but communicative competence (Hymes, 1972), as well as the processes involved in language acquisition which were thought to fine-tune performance in the negotiation of meaning during communication. **Communicative Language Teaching** (CLT; e.g., Brumfit & Johnson, 1979; Widdowson, 1978) evolved in concert with these changing views of language as an outgrowth of the situational approach to language teaching (Hornby, 1950) that had developed earlier in Britain (Richards & Rodgers, 2014, p. 46). Communicative Language Teaching was developed in connection with SLA theory that differentiated conscious and deliberate *learning* of language, as the result of explicit instruction and application of learning strategies, from unconscious *acquisition* of language, as a natural consequence of immersion and interaction in communicative contexts (Krashen, 1982, 1985). From this perspective,

a main function of instruction is to provide learners opportunities for real or realistic (e.g., simulated) communication opportunities, so acquisition can take place implicitly through the learner's cognitive mechanisms.

CLT approaches, such as the input-heavy **Natural Approach** (Krashen & Terrell, 1983; Terrell, 1977, 1982) or the interactive group and pair activities (e.g., Savignon, 1983) that became standard in L2 teaching, were widely adopted for teaching English in bilingual education programs, such as for Spanish-English bilinguals in California and French-English bilinguals in Canada, and for teaching English as a second language (ESL) within English-dominant countries, even while traditional formal approaches were maintained to a greater or lesser extent in other teaching contexts. In the attempt to encourage natural acquisition processes and student engagement and risk-taking in using the second language and interactively negotiating meaning, the focus of classroom teaching moved away from accuracy and correct performance and so away from pronunciation. The lack of attention to pronunciation was common in teaching within both the second language orientation of ESL programs in English-dominant countries and the bilingual orientation of programs aimed at mother tongue maintenance and strong competence in two languages that were current in both English-dominant countries such as the United States and dual-language countries such as Canada. At the same time, "visible speech" technologies were becoming popular for individualized use and tutoring outside of classrooms, such as in computer-enhanced language laboratories (Pennington, 1989; Pennington & Esling, 1996).

1990s/2000s: Fine-Tuning Communicative Tasks

After critiques of bilingual programs in Canada (see Genesee, 1987, for review) led to a reconsideration within second language acquisition (SLA) theory of the role of explicit feedback in fine-tuning learners' competence in a second language, L2 teaching in the 1990s and 2000s reinstated a focus on form that incorporated explicit attention to pronunciation within an overall goal of effective communication (as Pica, 1984, had earlier suggested). During this period, language teaching from a second language orientation was increasingly directed to learning language for a

specific field or type of job, as developed in **English for Specific Purposes** (ESP) methodology based on needs analysis and practical skills (Dudley-Evans & St. John, 1998), and/or for the completion of specific kinds of tasks, as developed in **Task-based Language Teaching** (TBLT; Ellis, 2003; Long, 2015). Both ESP and TBLT, and communicative methodology more generally, were adapted to accommodate a focus on form as an aspect of communicative performance. In addition, during this period, many new computer-assisted learning technologies were developed with a focus on pronunciation (for a review, see Chun, 2007, especially pp. 284–293).

However, those coming from other traditions, such as bilingual education in the United States (Ramsey, 2012) and multilingualism–multiculturalism in Europe and worldwide (Edwards, 1994) often maintained a content-based instructional orientation, which in Europe was the **Content and Language Integrated Learning** (CLIL) approach promoted to support multilingualism (Dalton-Puffer, 2007). Within these bilingual and multilingual approaches, pronunciation has not been regarded as a main concern of instruction, based on a view of accented speech as a normal fact of life in contexts of bilingualism and multilingualism that does not necessarily impede communication. Within this understanding of pronunciation, an English as an International Language (EIL) or **English as a Lingua Franca** (ELF) approach to teaching pronunciation emerged, with a goal of ensuring intelligible speech for encounters among speakers for whom English is the common language by focusing on a simplified core of communicatively essential pronunciation features (Jenkins, 2002).

2010s: Plurilingual Emphasis on Speaker Agency and Identity

Consistent with bilingual and multilingual, as opposed to second language, orientations to language learning and teaching, pronunciation had not previously been a main concern for the Language Policy Division of the Council of Europe's (2001) Common European Framework of Reference (CEFR) for Languages (e.g., in their scales for Spoken Language

Use and Phonological Control). However, recent, substantial revision of the pronunciation descriptors in the CEFR 2017 version (as reviewed in Chap. 5) reflect growing recognition of the relevance of a plurilingual approach to language use in Europe. Pronunciation has recently been reconsidered by Pennington (2015a, 2015b) from a perspective that can be seen as falling within the Council of Europe goal of *plurilingualism/pluricompetence*. As elaborated by Coste, Moore, and Zarate (2009), this goal includes being able to use several languages to varying degrees and for distinct purposes in intercultural interaction, and developing “linguistic tolerance,” that is, appreciation and respect for speakers with different language and cultural backgrounds. Pennington (2015a, 2015b) links these goals to the practices of *translanguaging* across two or more languages to create new meaning and express different aspects of an identity based in more than one language and culture (García, 2009; Li Wei, 2011), and to speakers’ use of pronunciation features from more than one dialect or variety for these same kinds of creative and expressive purposes.

A speaker might create or display meaning, identity, and communicative competence through use of pronunciation features from different languages, dialects, or varieties, such as in style-shifting (Eckert, 2000; Eckert & Rickford, 2001), performing double-voicing and mock-standard accents in crossing (Rampton, 1995), displaying intragroup and intergroup affiliations (Giles, Bourhis, & Taylor, 1977), or expressing a complex multilingual or metrolingual identity or repertoire (Otsuji & Pennycook, 2010). In these different ways, speakers are able to use features of language, including both segmental and prosodic features, to enact a linguistic competence that spans more than one speech community and linguistic system and to display their intentional “stancetaking” (Johnstone, 2009), through which they “create and signal relationships with the propositions they utter and with the people they interact with” (p. 31). A recognition of this fact breaks with the tradition of pronunciation competence as a feature of language outside the speaker’s awareness and control and based in only one language or variety and moves in the current era beyond second language phonology and towards a new concept of *multiphonology* or *pluriphonology* (Pennington, 2015a, 2015b) that is part of more social constructions of language learning and use (Block, 2003; Ortega, 2013).

Concerns in Designing Pronunciation Instruction

As our brief historical review demonstrates, pronunciation teaching has historically gone in and out of fashion: at times being marginalized and at other times being considered a central goal of language teaching. For instance, within the audiolingual approach which predominated in the 1960s, a great deal of concern was given to developing pronunciation accuracy through learners' repetition of native speaker recorded speech, whereas during the heyday of CLT in the 1970s and 1980s, pronunciation was sidelined. Currently, we see a resurgence of interest in pronunciation teaching in the "focus on form" emphasis that is now attached to CLT, TBLT, and ESP teaching goals and that exploits the vast array of technologies available for language teaching.

Yet knowledge about the nature of pronunciation and pronunciation teaching is generally limited and often reflects outdated and misguided notions. In the usual case, both teachers and students have a restricted conception of pronunciation, typically oriented to accuracy or correctness based on an implicit native speaker model, that limits their view of what is to be learned and taught in this area of language. In addition, their experience as language learners and teachers, and as communicators in their home speech community and in other speech communities, will have shaped their attitudes and beliefs about this important and highly visible aspect of communication. In many cases, teachers and learners will need to have their consciousness raised and to be reeducated in many areas relating to pronunciation as an aspect of language and communicative competence and as a component of the language class, in order for pronunciation instruction to be most effective and to achieve the best outcomes in relation to other areas of the language curriculum.

Despite the value many learners put on pronunciation proficiency (Derwing & Rossiter, 2002), language teachers do not always agree with their students on its importance, and even those who do can be daunted by the prospect of teaching pronunciation, so that pronuncia-

tion often remains a relatively marginalized area of the language teaching curriculum (Gilbert, 2012). Lack of time in a packed syllabus or lack of confidence in their own pronunciation ability or subject knowledge may be legitimate concerns for many English language teachers. Concerns may also stem from uncertainty about what areas of pronunciation to prioritize, what variety of English to use as a model—such as **Received Pronunciation** (RP), General American (GA), or a local variety of English—or what goal is appropriate or desirable in order to achieve mutual intelligibility among whatever groups learners desire or need to communicate with. In many cases, the teacher will need to consider pronunciation targets and communication goals aimed for a range of L1 and L2 speakers of English (Rogerson-Revell, 2011). Such concerns and considerations have gathered momentum in recent years, not only because of the natural evolution of the English language in different parts of the world, resulting in a wide range of accents and language varieties, but also because of the rapid and continuing spread in the use of English as a lingua franca between speakers with different first languages (Seidlhofer, 2011) that may be pressuring to an opposite trend towards a more universal variety. As noted by Ur (2012), these are part of natural *centrifugal* (dispersive) and *centripetal* (consolidating) forces of language.

Determining Pronunciation Models and Goals

It is obviously important for learners to have clear models, achievable goals, and realistic targets for pronunciation; but it is not always easy for the teacher to decide exactly what these should be. For example, should the teacher choose between an established, or “standard,” L1 model such as GB or GA, or is a local standard accent more appropriate? Similarly, should the learners be aiming to approximate native proficiency as closely as possible, or is a more restricted target such as “comfortable intelligibility” (Kenworthy, 1989) more achievable? Moreover, the teacher’s own priorities may not be reflected in the demands of the school or syllabus, nor of parents or the learners themselves.

Models

Many people believe they do not have an accent or that it is possible to speak without an accent. As we have seen, from a linguistic standpoint, all speakers have an accent, although some accents may seem less noticeable or marked than others due to their widespread usage or currency, and some will be considered more standard or acceptable than others in certain contexts or for certain purposes. A great deal of research has been conducted into listeners' attitudes and reactions to standard and non-standard or foreign-accented speech (e.g., Beinhoff, 2013; Bradac, 1990; Cargile, Giles, Ryan, & Bradac, 1994; Giles, 1970; Giles & Sassoon, 1983; Honey, 1989). A common finding is that listeners typically employ stereotypes based on accent to assign speaker attributes, such as hard-working or lazy, attractive or unattractive, friendly or unfriendly. As we saw in Chap. 1, attitudes to both L1 and L2 accents can be positive; for instance, a Scottish accent is often considered attractive by British and American listeners, and a French accent is often seen as sophisticated or charming. On the other hand, negative connotations or attributes can be associated with accents, such as German-accented English being considered as sounding formal or brusque, and with specific features of accent, such as English /r/ as pronounced by L1 German or Greek speakers (Beinhoff, 2013). Attitudes to accents can have significant consequences; for instance, positive reactions can help sell products or appease angry customers, while negative judgements can lead to discrimination, such as in job interviews or applications for asylum, with far-reaching consequences, such as limitations in employment, income, and mobility (Munro, 2003).

Traditionally, teachers have sought to use an L1 standard accent as the model for pronunciation teaching. Typically, this has been either GA for American English or RP for British English, depending on the learners' context and which variety is considered most useful to them. One of the main arguments for using such models is that they have been recognized as prestige varieties or accents and so using such models gives learners access to the power and prestige associated with that accent. A further argument is that a great many resources, including teaching resources such as textbooks and dictionaries, are based on these L1 accents. However, in the case of RP, Trudgill (1974) spec-

ulated that this accent was used only by a small minority of L1 English speakers, which he put at approximately 3% of the British population, and others have noted that an RP accent is increasingly considered old-fashioned and elitist (Cruttenden, 2014; Jenkins, 2000). As Wells (1997) observed over twenty years ago, the RP accent, which is often considered as an idealized form of “proper English,” if considered in real terms is changing to incorporate many features that are common in other British varieties of English. Thus, the concept of RP English is perhaps as much myth as reality, and so attempting to have students emulate this accent is unrealistic. In the case of L1 accents more generally, it can be argued that trying to make learners adopt an “alien” accent involves a threat or loss of their own L1 identity and is morally wrong (Porter & Garvin, 1989).

Recognition of the links between accent, identity, and social status has led many teachers and researchers to question the choice of pronunciation model for L2 learners. Many scholars in the United Kingdom have suggested replacing the term “RP” with more inclusive alternatives such as “BBC pronunciation,” “Non-Regional Pronunciation” (Collins & Mees, 2013, p. 4), or “Standard Southern British English” (SSBE). The term used in this book for a model of British English is General British (Cruttenden, 2014), which represents a form of standard English which is not easily regionally identifiable and is used widely by educated speakers across the British Isles. Similar standard varieties, such as General Australian and General American (also sometimes referred to as “Middle American,” reflecting what is viewed as the less distinctly accented speech of the interior as contrasted with the coastal and border regions of the United States), have been used as models by millions of students learning English as L2 around the world. Choosing such a model does not suggest that this is the “best” or “correct” accent (although students sometimes assume this); but it can be argued that such models provide a clear reference point, particularly for beginner students. Whatever accent is chosen as a model, it is important to bear in mind that there is no single version of an accent, whether GA, GB, or some other standard. For every form of English, there is a range of versions, depending, for instance, on a speaker’s age or social class; and all accents evolve over time.

One view which is gaining currency is that it is neither necessary nor desirable to provide a single pronunciation model for students, but rather multiple models can be taught (e.g., Carrie, 2017; Esling, 1987; Mompean, 2004, 2008). Mompean (2004), for example, reviews the options for choice of pronunciation models for teaching English in Spanish universities. He suggests the value of having both reference and comparison pronunciation models, with the former used for systematic description and speaking/listening practice and the latter used only for description and listening practice. He considers the suitability of RP and other English accents for Spanish L1 learners of English based on four criteria: (i) admiration by students and other stakeholders for the model accent; (ii) specific need for and use of the model accent; (iii) attitude/tradition of the administration of the educational unit regarding the model accent; and (iv) availability of the model accent (e.g., in terms of teaching materials). On the basis of data gathered from students and university documents, Mompean (2004) proposes that the British RP accent seems most appropriate as the reference model and an educated American variety such as GA seems most appropriate as the primary comparison model for English language university students in Spain, with other accents, such as Irish English, being used for some purposes of comparison. This work by Mompean to decide teaching models for pronunciation illustrates teaching practice based on a practitioner's research into student preferences and their specific educational context.

It has been suggested (e.g., Esling, 1987; Levis, 2016; Pennington, 1996) that learners should be aware of the range and flexibility in the production of sounds (e.g., that the vowel in the word *man* can be pronounced [æ], [ɛ], [a], or [ɑ]), and that they will learn more effectively from exposure to a wide variety of contextual variants, as discussed further in Chap. 4, including through implicit learning mechanisms (see Chap. 2). Esling (1987) advocated teaching pronunciation in relation to a range of models, focusing pronunciation instruction on “collection not correction” (p. 469), and Pennington (1996) championed a “variationist philosophy” in which “[a] major part of the language teacher's job [is] one of providing the students with a broad range of experiences within the language and a diversity and quantity of input in the way of speech samples on which to base their own phonology” (pp. 17–18). A variety of

pronunciation exemplars are applied in what is known as **high variability perceptual** (or **phonetic**) **training** (HVPT), as first tested by Logan, Lively, and Pisoni (1991) and as described in more detail in Chap. 4.¹

The most available model for most learners is their classroom teacher, who, in many cases, will not speak with an L1 or standard accent. A teacher who is an L1 English speaker may, for example, have a regional British, American, or Australian accent. Teachers who speak with a regional accent need to be aware of how their own accent deviates from the standard model and to consider which of the two to choose as the main teaching model. A teacher who is an L2 English speaker will similarly need to decide if an L1 or L2 standard model (e.g., Singaporean or Nigerian English) is more suited to learners' needs (see further discussion in Chap. 4). In addition to the teacher's own model, films, television, the internet, and social media give easy access to a wide variety of both L1 and L2 speakers, including well-known people such as politicians or celebrities who might prove appropriate and aspirational models for language learners, particularly those who share the learners' L1 background.

Increasingly, it is becoming recognized that L2 English teachers who have achieved a high level of proficiency in English can provide an appropriate model for their students and be in a strong position to teach English, having learned the language themselves (Pennington & Richards, 2016). Indeed, Levis and Grant's (2003) study of two ESL instructors, one an L1 English speaker and the other a highly proficient bilingual speaker, concluded that both were equally effective pronunciation teachers.

Debates about choice of pronunciation model have been ongoing for many decades (see Honey, 1989; Widdowson, 1994) and no doubt will continue for some time to come, given the economic, academic, and social value of knowing one or another variety of a language. However, the growing use of English as an international language or lingua franca has fueled the discussion regarding choice of pronunciation model in recent years. Much of this discussion reflects the need for "international intelligibility," increasingly between speakers with different L1s who are using English as a common L2. At the same time, there is obviously practical and social value in being able to speak the variety or varieties of English used in the local speech communities and communities of practice where one works and lives.

The discussion of pronunciation models can be seen in a more positive light if viewed in terms of acquiring an additional accent in order to facilitate mutual phonological intelligibility, rather than trying to eliminate or reduce an undesirable L1 accent, that is, so-called “**accent reduction**.” The social significance of pronunciation, and accent in particular, has led to the provision of commercial accent reduction services to help concerned individuals “neutralize” their accent. Accent reduction has become big business in many parts of the world where companies charge a considerable amount of money to reduce foreign or regional accents. In the United Kingdom, private providers work in companies and schools, for instance, giving primary children elocution lessons to help them correct accents perceived as “common” or uneducated. While some such providers may have the necessary skills and knowledge to improve intelligibility in some cases, the aims and motivation behind the industry appear to be ethically questionable.

The concept of **accent addition** has been opposed to “accent reduction” in order to reflect the aim of expanding an individual’s phonological repertoire, instead of viewing the L2 accent as a threat to the speaker’s L1 accent and sense of identity (Jenkins, 2000). A view of accent in terms of dual or multiple competences, in which users adjust their pronunciation according to whether they want to express their local or global selves, or various stances and styles existing within global and local speech communities, may allow learners “the freedom to choose one’s belongingness” (Shaw, 2008, p. 49). It may also help remove the threatening, competitive element associated with acquiring an L2 accent, if both L1 and L2 accents are equally valued, and consequently the choice of model becomes less important. This positive view of language varieties is parallel to that in sociolinguistics which values nonstandard (e.g., ethnic or regional) speech (e.g., Eckert, 2000; Eckert & Rickford, 2001; Fought, 2006; Labov, 1966, 1972).

Goals

When considering what to teach in pronunciation instruction, there is sometimes a confusion between the concepts of model and goal. A pronunciation model is a set of pronunciation forms representing a particular

language variety or accent which can be used as a reference point against which to measure pronunciation accuracy or appropriateness. Thus, a pronunciation model provides examples of how sounds and words are normally pronounced in a particular variety; for example, the sound /ɑ:/ is the normal vowel phoneme in the word *bath* and *half* in GB while /æ/ is the normal phoneme for these words in GA.

The pronunciation goal is the learning target that the teacher or learner sets in terms of the level of proficiency which a pronunciation student will have to achieve in order to communicate effectively. The goal may therefore vary depending on the learner's context or needs. For example, learners who will be dealing mainly with L1 English speakers might aim for a target of near-native proficiency. That person might set a goal to eliminate or reduce segmental and prosodic features of the L1 accent that cause misunderstanding, aiming for an accent that is less identifiable in terms of the L1, thus more "neutral," as has been recommended for international call center work (Chap. 7). However, learners who intend to interact mainly with other L2 speakers may be able to communicate effectively with a more limited degree of phonological competence, or with a distinctive L1 accent, as long as they have a sufficient mastery of core phonological features, as suggested by some proponents of ELF, to ensure intelligibility between speaker and hearer. In this case, the learning goal for pronunciation may be a more limited or focused one centered, for example, on intelligibility rather than "full competence" (Ur, 2012). An instrumental-pragmatic goal of acquiring English for limited lingua franca communication suggests different priorities for instruction than a social-integrative goal of developing an international or Global English identity (see discussion in Chap. 2).

Intelligibility as a Goal

Levis (2005) introduced two contrasting principles as goals for teaching pronunciation, one based on nativeness and the other on intelligibility. According to the **Nativeness Principle**, the goal in L2 learning is to develop speech that matches that of a native speaker model. According to the **Intelligibility Principle**, the goal in L2 learning is intelligibility, no matter

how much the speaker's accent diverges from that of a native speaker. Isaacs and Trofimovich (2012) maintain that “[f]ew L2 researchers and practitioners would disagree that intelligibility is the appropriate goal for L2 pronunciation instruction” (p. 477), and Levis (2016) has recently made the point that a focus on intelligibility can avoid spending “precious classroom time on something that is not likely to pay off” (p. 4).

Intelligibility is increasingly seen as an appropriate and achievable goal for many learners, and recent research in the field continues to support this view (Sewell, 2013; Sung, 2013). As we saw in Chap. 1, intelligibility can be understood as an interaction between a speaker and a hearer in terms of the ability to produce and to recognize words and utterances (Smith & Nelson, 1985), and to speak with sufficient fluency that continuity in production and perception does not get disrupted. A further distinction has been made between intelligibility and comprehensibility, with the former seen as “the extent to which a speaker's utterance is actually understood” (Munro, Derwing, & Morton, 2006, p. 112) and the latter seen as “the ease or difficulty a listener experiences in understanding an utterance” (Derwing & Munro, 2015, p. 5).

Although most educators would agree that intelligibility is an important goal for language learning and teaching, they may place different emphases on what a goal of intelligibility implies. For example, the goal of “comfortable intelligibility” (Kenworthy, 1989) implies reaching a level of intelligibility where both listener and speaker can communicate effectively without undue effort or discomfort. The concept of “international intelligibility” (Jenkins, 2000) raises the question of “intelligible to whom?” While most earlier discussions of intelligibility have assumed being understood by an L1 listener, the case has increasingly been made to consider understanding from the point of view of both L1 and L2 listeners and, in many ELF contexts, to consider mutual intelligibility between L2 speakers and listeners. Research suggests that “many so-called NSs [native speakers] can be far less intelligible in global settings than well-educated proficient speakers of a second language” (Moussu & Llurda, 2008, p. 318).

Despite increasing recognition of the relevance of “international intelligibility” as a realistic goal for many language learners, research suggests that many L2 English teachers and learners still strongly adhere to L1

English pronunciation norms and prefer to aim for native or near-native competence (Jenkins, 2005; Kuo, 2006; Scales, Wennerstrom, Richards, & Wu, 2006). Research shows (e.g., as reviewed in Beinhoff, 2013, pp. 29–32) that L2 teachers and learners, though they may be “attracted by the concept of ELF, on the other hand, ... regard NNS [non-native speaker]-accents as stigmatized and prefer to stick to NS [native speaker]-accents when it comes to defining their aims for learning and teaching English...” (p. 31). For pronunciation teachers, the desire to reach a high level of target language proficiency is understandable. As with all language teachers, pronunciation teachers need to have at least receptive competence in one or more standard varieties of the language they teach, whether it is their first language or not. The fact that many learners also aim for near-native competence in pronunciation suggests that factors other than intelligibility, such as concerns relating to performance, proficiency, and social acceptance, also come into play (Gluszek & Dovidio, 2010; Kuo, 2006; Timmis, 2002). Rather than trying to convince learners of the lack of importance of achieving a native-level accent, teachers might perhaps focus on the fact that a high level of proficiency and performance, in terms of intelligibility and fluency, and of social acceptance, can be achieved regardless of accent.

Typically, accent and intelligibility are seen as closely related, and for many decades pronunciation teaching has aimed at achieving, as far as possible, a native-like accent or at least reducing a foreign accent. Even today, the goal of unaccented speech or “accent reduction” remains central to many learners’ and teachers’ view of second language pronunciation. There are two issues with this goal. The first, mentioned earlier, is that the deficit concept of “accent reduction” is increasingly being replaced with the notion of “accent addition,” whereby learners aim to add an additional accent to their linguistic repertoire. This more constructive orientation avoids the requirement of learners to lose their L1 accent, and for some, the implication of losing their L1 identity. The second issue is that accent and intelligibility, while sometimes closely related, are not completely interdependent. This fact has been illustrated by research such as Munro and Derwing’s (1995) study of Mandarin speakers of English, where all the speakers whose speech was evaluated as fully intelligible (according to listeners’ ability to produce totally correct

transcriptions of their speech) were assessed on various points of a 9-point scale for accentedness. As documented in that study, speakers can sometimes be judged 100% intelligible and yet have strongly accented speech.

The growing emphasis on intelligibility as a goal in pronunciation teaching rather than near-native or nativelike competence has been reinforced by the increasing use of English as a lingua franca among L2 speakers. It has also been fueled by empirical research such as that of Munro and Derwing (1995) providing evidence that accent and intelligibility are partially independent, and by studies showing that few adult learners achieve nativelike pronunciation in a second language (Flege, Munro, & MacKay, 1995).

Establishing a Phonological Core for Intelligibility

There have been several attempts to establish a common core of phonological features that would be essential for mutual intelligibility. One of the earliest was by Hockett (1958), who aimed to find the core phonological features needed to ensure intelligibility among L1 dialects of English. Later research by Jenner (1989) took a similar approach, attempting to determine “what all native speakers of all native varieties have in common which enables them to communicate effectively with native speakers of varieties other than their own” (p. 2, emphasis in original). Unlike Hockett, Jenner had the pedagogical aim of uncovering a set of pronunciation teaching priorities for L2 learners of English which “would offer the learner a guarantee of intelligibility and acceptance anywhere in the world” (p. 2).

Up to this point, the focus was on intelligibility of non-native speakers of English by native speakers of English, particularly in “expanding circle” contexts (Kachru, 1985) where English is generally used as a foreign, rather than as a first or a second, language. Since then, however, the use of English has changed considerably, including the predominant use of English in expanding circles now between L2 speakers (Graddol, 2006; Ur, 2012). Some researchers have considered how mutual intelligibility can be sustained in such contexts where English is primarily used as a lingua franca. Jenkins (2000, p. 95) built on Jenner’s work, suggesting

the need for “some sort of international core of phonological intelligibility,” which she refers to as the “Lingua Franca Core” (LFC). She sees this as “a set of unifying features which at the very least, has the potential to guarantee that pronunciation will not impede successful communication in EIL settings” (ibid.).

Jenkins proposes that this core of features should be taught internationally to ensure intelligibility across varieties of English, claiming that it would provide more realistic and achievable classroom teaching priorities than teaching the complete repertoire of native speaker phonology. This core then would prioritize those pronunciation features which are key to mutual intelligibility in EIL or ELF contexts. Jenkins (2000) claims that such a core would provide more realistic and achievable classroom teaching priorities and would “scale down the phonological task for the majority of learners by leaving to the individual learners’ discretion and to later acquisition outside the classroom the learning of peripheral details, and focusing pedagogical attention on those items which are essential in terms of intelligible pronunciation” (p. 123). The essential features of the Lingua Franca Core are given in Fig. 3.1. A summary of the phonological features which Jenkins considers nonessential or outside the Lingua Franca Core are given in Fig. 3.2.

Jenkins points out that such a restricted repertoire is only relevant in contexts where English is being used as a lingua franca between L2 speakers and not in situations where communication occurs between L2 and L1 English speakers. Cruttenden (2008), similarly, considers using a restricted phonological repertoire for the learning and teaching of English as an additional language.² He introduced the terms “**Amalgum English**” and “**International English**” in the seventh edition of *Gimson’s Pronunciation of English* (Cruttenden, 2008) in view of the changes that English has undergone since the first edition appeared in 1962. *Amalgum English* refers to a language goal considered appropriate for learners who use English “as an L2 and/or lingua franca within their own country (and maybe including neighbouring countries) and who may only have limited meetings with L1 [English] speakers” (Cruttenden, 2008, p. 317). *International English* refers to a language goal for speakers who use English “as a lingua franca on a more international basis and need a minimum standard for occasional communication (eg non-English

1. Consonants

- All consonants, except for /θ/ and /ð/ sounds, as in *thin* and *this*.
- Allophonic variation is acceptable as long as it does not cause confusion with another phoneme (e.g., Spanish pronunciation of /v/ as [β] in word initial position can be heard as /b/, e.g., *vowels* is heard as “bowels”).
- Rhotic /r/ rather than non-rhotic varieties.
- British English /t/ between vowels, rather than American English flapped [ɾ] (e.g., *butter*, *water*).

2. Additional phonemic requirements

- Aspiration of word initial voiceless stops (e.g., *pin*, *top*, *cat*).
- Shortening of vowel sounds before voiceless consonants (e.g., *sat* vs. *sad*; *seat* vs. *seed*).

3. Consonant clusters

- No omission of sounds in word initial clusters (e.g., *promise*; *string*).
- Omission in middle or end of words is only permissible following L1 English rules (e.g., *factsheet* can be pronounced “facsheet” but not “fatsheet” or “facteet”).
- *nt* between vowels should follow British rather than American English pronunciation (e.g., *winter* should be /ˈwɪntər/ rather than /ˈwɪnər/).
- Vowel insertion is permissible rather than vowel deletion (e.g., *product* can be /pəˈɒdʌktɒ/ but not /ˈpɒdʌk/).

4. Vowels

- Vowel length contrasts should be maintained (e.g., in *live* vs. *leave*).
- L2 regional qualities are acceptable as long as they are consistent, except substitutions for /ɜ:/ (e.g., /bɜ:d/) which cause confusion.

5. Stress

- Appropriate use of contrastive stress (e.g., *my sister wears glasses*), with stress on the final word, as a neutral, or unmarked statement of fact, vs. *my sister wears glasses*, in the marked contrastive meaning which emphasizes that it is my sister and not someone else who wears glasses).

Fig. 3.1 Essential features of Jenkins’ Lingua Franca Core (adapted from Jenkins, 2002, pp. 96–97)

- The consonant sounds /θ/ and /ð/ and the allophone [ʔ].
- Vowel quality (e.g., the difference between /bʌs/ and /bus/ as long as quality is used consistently).
- Weak forms of function words (e.g., *to* /tə/, *of* /əv/, and *from* /frəm/). The use of schwa can hinder intelligibility more than full forms in EIL contexts.
- Other features of connected speech, especially assimilation (e.g., assimilation of final /n/ in *green* to the bilabial position of /p/ in *pen* in the phrase *green pen* pronounced as “greem pen” (Jenkins, 2002, p. 149) or elision of the first vowel in *perhaps* when spoken casually).
- Placement of word stress, which varies considerably across L1 varieties.
- Direction of pitch movement, whether to signal attitude or grammatical meaning.
- Stress-timed rhythm.

Fig. 3.2 Nonessential features of the Lingua Franca Core (Jenkins, 2002, p. 98)

speaking businessmen who use English as the common language between them)” (ibid.). Cruttenden’s summary of the core pronunciation features of Amalgum English and International English are given in Figs. 3.3 and 3.4.

There are obvious similarities as well as notable differences between Cruttenden’s and Jenkins’ proposed phonological cores for English in international lingua franca contexts. In terms of similarities, both emphasize the need to give greater priority to mastery of the consonant system than the vowel system. Cruttenden (2008) claims that “vowel contrasts in general appear to be less crucial to intelligibility than consonant contrasts so that a major simplification of the vowel system is possible for International English” (p. 331). Both researchers emphasize length rather than quality differences in vowels, and both see the use of schwa as non-essential to a phonological core (even though it is a core component of English stress and rhythm). Both restrict the focus on suprasegmental features to stress, either contrastive stress (Jenkins) or stress in multisyllabic words (Cruttenden), dispensing with tone or intonation in the sense of pitch movement, while acknowledging its role in conveying the subtleties of communicative meaning. It seems that tone or intonation, while essential for fully intelligible and meaningful ELF and EIL communication, is seen as a feature that does

- (1) General aim: easy intelligibility by native speakers.
- (2) Consonants:
- i. Insist on aspirated plosives but allow dental or retroflex /t, d/ and palatal /k, g/.
 - ii. Insist on /f, v, s, z/ but allow conflation of /ʃ, ʒ/ and /θ, ð/. /h/ required but allow velar/uvular replacements.
 - iii. Insist on /tʃ, dʒ/ distinct from /tr, dr/.
 - iv. Allow any variety of //l/. Allow prepausal and pre-consonantal /r/ and /r/ = [ɾ]. Allow insertion of /g/ following /ŋ/. Discourage /w/ = [ʋ].
 - v. Insist on consonantal clusters (apart from usual reductions allowable in RP).
- (3) Vowels: a possible reduction to:
- i. Short vowels /ɪ, e, æ, ʊ, ə/
 - ii. Long vowels /i:, e:, a:, ɔ:, ɜ:, o:, u:/
 - iii. Diphthongs /aɪ, aʊ, (ɔɪ)/
- (4) Connected speech:
- i. Insist on nucleus movement and basic tunes

Fig. 3.3 Summary of “Amalgum English” phonological core (Cruttenden, 2008, p. 329)

not require attention to teaching. They also both emphasize that such a core would ensure “minimal” intelligibility and should not be seen as an ultimate goal for learners who want to acquire a higher level of proficiency. However, it can be noted (as reviewed in Chap. 2) that while children learning their L1 start out with a reduced set of phonological contrasts which they gradually elaborate into a full set of contrasts including contextual (positional and social) variants, it is less likely for an adult’s phonological system to develop in this way. When adult learners start out with a reduced set of phonological contrasts, as a result of simplified or highly restricted input (e.g., in pidginization) or as a result of the combined influence of L1 transfer and language universals in L2 learning, so that they fail to differentiate all the sounds of the L2, the reduced set can become entrenched and difficult to alter. It is also possible that explicit instruction in the LFC will entrench that phonologically reduced core, and the idea of elaborating or adding to it later may present significant difficulties.

- (1) General aim: minimal intelligibility in the use of English in international lingua franca situations.
- (2) Consonants:
- i. Allow voicing distinctions to be made using different features than those used by native speakers.
 - ii. All forms of /r/ and /l/ are allowed but distinction between the two to be given high priority (even for those speakers from Asia who find it difficult, e.g., Japanese and some Chinese). As for Amalgam English /r/ should follow the spelling and any sort of /r/ allowed.
 - iii. Distinction between /v/ and /w/ should be insisted on; use of /ʊ/ for either or both discouraged.
- (3) Vowels:
- i. A reduction in the vowel inventory to five short and five long vowels is allowable (it will be used naturally by many learners, e.g., Bantu speakers).
- (4) Connected speech:
- i. Some attempt should be made to place the accent on the usual syllable of polysyllabic words but no attempt need be made to use the weak forms of English or the weak syllables in polysyllabic words (i.e., no reduction to /ə/ need be made).
 - ii. No effort need be made to learn native intonation patterns of L1 English.

Fig. 3.4 Summary of “International English” phonological core (Cruttenden, 2008, p. 333)

In terms of differences, Cruttenden (2008) includes more specific points related to consonants than Jenkins (2000) does. For “Amalgam English” he suggests “allow[ing] conflation of /ʃ, ʒ/ and /θ, ð/” but “insist[ing] on /tʃ, dʒ/ [remaining] distinct from /tr, dr/,” and for International English he maintains that the “distinction between /w/ and /v/ should be insisted on.” The rationale for the specific guidelines for phonological core varieties is unclear in terms of criteria based strictly on intelligibility, as there would not seem to be a greater possibility of lexical confusion involving the contrasts of /tʃ, tr/ or /dʒ, dr/, or even of /w, v/, in context than the contrasts of /ʃ, ʒ/ and /θ, ð/. In general, the rationale for determining essential versus non-essential aspects or “peripheral details” (Jenkins, 2000, p. 23) is not clear.

The specific proposals made by Jenkins (2000) and Cruttenden (2008) seem to be based (implicitly) on the norms of a British variety (e.g., the

British vowel system) as the source for acceptable forms and allowable simplifications or divergences, such as Jenkins' statement that "L2 regional qualities are acceptable as long as they are consistent, except substitutions for /ɜ:/ (e.g., /bɜ:d/), which causes confusion." This one exception to an allowance for "L2 regional qualities" suggests an underpinning non-rhotic British norm, although the LFC otherwise seems to be based on a rhotic variety of English.

In spite of issues that can be raised about these reduced phonological inventories and their specific details, the increasingly dominant use of English as an international language by L2 speakers has fueled and spread awareness of the need to consider what appropriate goals are for L2 learners. The rationale for teaching from an ELF perspective was given by Jenkins (2012) in an article in *English Language Teaching Journal*, where she makes the observation:

Despite the phenomenal increase in the use of ELF around the world, the prevailing orientation in English language teaching and testing, and ELT materials remains undoubtedly towards [English as a native language], with correctness and appropriateness still widely driven by [native English speaker] use regardless of learners' current or potential communication contexts.... Thus, learners of English who are more likely to use their English to communicate with other [non-native English speakers] than with [native English speakers], more often than not with no [native English speakers] present, are still being encouraged to aim for the kind of English that British or North American English speakers use among themselves. And when students around the world have completed their English language courses, it is this same native English (again, typically British or North American) that is assessed in the supposedly 'international' ELT examinations. (p. 487)

While we are sympathetic to the rationale behind the phonological core approach, the notion of learning and teaching according to a reduced set of phonological features has prompted considerable discussion among researchers and teachers. Some scholars are critical of the concept of ELF more generally, suggesting that although advocates of ELF such as Jenkins recognize diversity of varieties, there has nonetheless been a tendency by some to view ELF as a discrete language variety

which can be used as a teaching model. For instance, Beinhoff (2013, p. 13) notes how many different L2-influenced forms occur in ELF communication, doubting that there is, or can be, a general ELF variety. Considerable critical discussion has ensued, as reflected in Sewell's article in response to Jenkins (2012) article. Sewell (2013) argues that:

... in some ways, neither ELF nor its opponents have come to terms with the complexities of English in a globalized world. By defining ELF according to how it differs from native-speaker language use, ELF researchers have tended to essentialize and exaggerate these differences, creating a false dichotomy between 'ELF' and 'non-ELF'. At the same time, the conservatism of language teaching and testing in many parts of the world also suggests that there is further scope for an enhanced awareness of language variation. A desirable outcome of the debate would be the realization that all language use—whether by native or non-native speakers—is variable, emergent, contextual, and subject to hybridity and change. (p. 8)

Nevertheless, Sewell (2013) seems somewhat ambivalent in his view of the LFC and does concede some of its practical benefits:

The message for learners seems to be that while ELF rightly emphasizes flexibility, maximizing this still requires the hard work of acquiring something resembling native-speaker competence. In addition to elucidating communicative strategies, ELF research may be able to help learners in this regard by identifying the features that are most important for international communication (as in the Lingua Franca Core developed by Jenkins *op. cit.*). This may not adequately address the social significance of language features, but it does free up time that can be used for more profitable learning activities. (p. 8)

From a teaching perspective, the debate has suggested that there is more support for such an approach from L1 English teachers (e.g., Walker, 2011) and less from L2 English teachers (e.g., Dauer, 2005; Kuo, 2006). Some teachers question the features of the core itself and others argue that it limits them unnecessarily and unrealistically, given the strong connections of pronunciation to identity and presentation of the self. In addition, phonologically simplified speech can be stigmatized (e.g., as showing lack of education) and so endorsing it can be challenged in terms of putting students in a marginalized status.

Yet it can be argued that a phonological core approach provides a useful reference point when deciding priorities for a pronunciation teaching syllabus, especially when time for teaching pronunciation is quite limited, and offers a sensible starting point for many learners of English, particularly those who will use English primarily in lingua franca contexts. Also, as Sewell (2009) concludes, “one advantage of combining the WE [World English] and ELF perspectives and adopting an intelligibility-oriented approach is that it may serve to counter the perception that local models hinder global intelligibility” (p. 42).

While we grant that the LFC approach has value in serving to reinforce the legitimacy of models of English that are not referenced to native speaker norms, teaching the LFC could mean that teachers give less emphasis to the social aspects of pronunciation, as Sewell (2013) points out. Such features, involving phonological variation, speaker positioning, and style shifting, have been recognized as central aspects of communicative competence and expression of identity, and, as we suggested in Chap. 1, are part of what can be defined as pronunciation competence. The important point to bear in mind is that the LFC approach represents a quite different orientation to pronunciation: one aiming to teach a basic level of skill sufficient for meaningful communication and intelligibility in restricted contexts, among speakers with different L1s. It seems especially appropriate when teaching and learning time and/or learners’ goals for English usage are limited. Such an approach is unlikely to be appropriate for those aiming to teach and learn language to an advanced level of skill and for many real-world communicative purposes (see further discussion in Chap. 7).

Deciding Teaching Priorities

As well as the issues involved in deciding what are appropriate models and goals for pronunciation teaching and learning, pedagogical concerns also often stem from uncertainty about what areas of pronunciation to prioritize. While learners are often aware that they have problems with pronunciation and may be keen to address these, they are unlikely to be able to

judge their relative severity or significance. It is therefore part of the teacher's job to be able to evaluate a learner's pronunciation as part of deciding which areas are priorities for teaching. In addition, decisions about teaching priorities must be taken in consideration of other factors such as learners' proficiency level, their end purpose for language learning, and what types of instruction can be carried out effectively and efficiently, and are likely to give the best results in the long run. Teaching priorities may be further constrained by limited time and resources, and by the necessity to follow a prescribed syllabus or curriculum. This has been the case in many countries where teachers have to follow a national curriculum which may not prioritize pronunciation or oral skills more generally. In this way, pronunciation teaching is similar to other areas of language teaching in which an important role for the teacher is deciding what and how to teach based on learners' needs and wants within their contexts of language use, in relation to teaching context, time, and constraints.

Learner Needs Analysis

A learner's needs will vary depending on factors such as the learner's entry level of pronunciation and purpose for learning, for instance, for occupational purposes or to pass an academic exam. Other factors need to be taken into account, including whether English will be used primarily for local or international communication, and consequently whether a local or international standard is more appropriate as the main pronunciation model, and what resources and time are available to be devoted to pronunciation instruction.

A key part of analyzing needs and deciding teaching priorities for pronunciation teaching is assessing learners' level of proficiency in perception and production of oral English. As regards receptive competence, several textbooks have diagnostic tests which help assess pronunciation perception (e.g., Gilbert, 2012; Rogerson & Gilbert, 1990). In order to assess oral production, a sample of each learner's speech can be collected, such as a recording of spontaneous speech, responses to specific prompts that offer a degree of control of the speaking task and output, and/or a written passage or word list read aloud (assuming that the learners have

adequate literacy in the L2). There are advantages and disadvantages to each of these methods of obtaining a speech sample. Spontaneous speech is more natural than the other options and might allow for greater fluency than either a controlled speaking task or reading a written passage; however, some speakers are more at ease speaking spontaneously than others, and the unplanned nature of the discourse makes it hard to compare across speakers. A controlled task such as picture description or story-telling based on a multiple-panel cartoon can elicit natural speech while also encouraging certain words and discourse structures to occur in the speech sample, thus allowing for greater comparability across speakers. A recording of a written text read aloud allows for easy comparison across speakers and facilitates the inclusion of specific pronunciation features in a text but can make some learners anxious and more prone to errors than usual (Hammerly, 1982), in addition to encouraging disengaged reading-prosody and spelling-pronunciation that diverge far from natural speech. Reading common words in a list is easier than reading a written passage aloud and will generally elicit a speaker's most careful pronunciation, and the list can be designed to incorporate multisyllabic words to give an idea of the speaker's prosodic competence. Ideally, a sample containing more than one type of speech is probably optimal to overcome the issues associated with each (Levis & Barriuso, 2012).

The issues involved in assessing speech samples are discussed in more detail in Chap. 6. However, it is useful from a teaching point of view to be aware that subjective listener evaluations of oral proficiency can be just as accurate and dependable as objective or acoustic measures, particularly for intelligibility, which is a perceptual phenomenon (Kibishi, Hirabayashi, & Nakagawa, 2014). On the other hand, teachers as language experts with considerable exposure to L2 speech may be more lenient or more able to figure out what learners are trying to say than non-expert judges. However pronunciation is assessed, the outcomes not only can help the teacher decide teaching priorities but, ideally, can also be used to create pronunciation profiles of individual students. Such profiles can help learner and teacher record progress and also provide a realistic evaluation for both those students who are over-confident of their ability and those who underestimate their pronunciation performance.

It is also important for the teacher to be respectful of the learner's own goals and priorities, particularly in situations in which the teacher's and student's goals do not coincide—for example, when a student wishes to aim for a target of “native” proficiency which the teacher believes is not realistic or appropriate to the learner's needs or the contexts in which the L2 is used. In such cases, the teacher may offer at least some material or lessons based on a native speaker model while presenting evidence that nativelike proficiency is for most learners neither achievable nor necessary, given the strong influence of L1 pronunciation norms, and has psychological and social consequences in terms of one's core identity as tied to the L1 language and culture (Pennington, 2018). While “near-native” proficiency is a more sensible goal and one that allows flexibility in how closely one would have to approximate native speaker norms, the teacher can point out that this goal is still a challenging one in terms of the time and focus one would have to put on pronunciation to achieve it and that it also has consequences for personal and social identity. The teacher might further point out that both “native” and “near-native” proficiency are a matter of perception which varies depending on the perceiver's experience, judgement, and the specific accent model that the person has in mind. Learners can moreover be reminded of the many factors that go into positive assessments of speakers and that a near-native or nativelike accent is not likely to be a key or deciding factor. Examples of favorite actors or sports stars can help underscore these points and also the fact that a person's accent may be one of the things others like about the person.

At the same time, in considering learner needs in the area of pronunciation, teachers must bear in mind that reactions to non-standard or “deviant” pronunciation seem to be stronger, especially among native speakers, than for other aspects of language. As Mey (1981) maintains, “on a scale of the native speakers' tolerance of linguistic deviance: they instinctively abhor phonological deviance, hate the morphological sort, merely dislike the syntactic, and can live with the semantic” (p. 69). What is considered deviant and what is considered appropriate in pronunciation evolves over time, as language—like everything else employed in human expression and display—exhibits shifting fads and fashions, resulting in concomitant shifts in tolerance and prejudice. However, there remains a common stigma associated with non-standard accents, both L2 pronunciation and nonmainstream (e.g., regional, ethnic, or minority) L1 varieties.

Prioritizing Pronunciation Teaching Areas

The learners' needs and motivations are the starting point for deciding both what and how to teach. In deciding teaching priorities, however, there are many other factors beyond individual learner characteristics and purposes that the teacher needs to consider in order to achieve the ultimate aim of helping learners communicate effectively. We have already seen that intelligibility is an essential component of effective communication but that other factors are involved. For instance, a learner can have intelligible pronunciation yet not be very comprehensible, requiring considerable effort to be understood for reasons other than pronunciation *per se*. A failure to communicate effectively or to be understood can stem from a variety of reasons, including choice of words and grammatical structures, as well as features of speech that have phonological effects at the level of prosody and articulation in connected speech, including lack of or incorrect grouping of words, and lack of fluency (inappropriate hesitation or pausing, or difficulty finding words). Problems in communication can also arise from a lack of attention to interpersonal dynamics and social meaning as conveyed by pronunciation, especially by prosody, in creating pragmatic or functional impact—for example, an L2 speaker's inability to impress a message on a listener because of ineffective use of stress and intonation. Awareness of such factors will help the teacher decide teaching priorities, but the teacher will also need to consider other criteria to be able to select specific elements of pronunciation to teach.

Learner Errors

An obvious criterion for deciding teaching priorities would seem to be learner errors. Despite ongoing debates about what constitutes an error, most teachers are confronted regularly with having to deal with L2 learner speech which contains features that deviate from the norms of the L1 or majority speech community and which, in some cases, impede intelligibility. In the example below, an L1 Spanish speaker has several pronunciation deviations which can be identified as substitution of L1 vowels

(apparently, spelling-pronunciations based on Spanish)—in *maybe*, *limit*, *main*, *bulk*, and *business*—in addition to two L1-influenced substitutions for the consonants in one word (*have*) that might not prevent a listener from understanding; other deviations are so severe that they make understanding impossible (as represented by ??; the two dots represent a short pause, four dots a longer pause):

'**maybe** [maɪbi] .. I don't [t] 'know .. and so .. the ?? **distri**'**bution** .. 'tries to
'**limit** [li.mi:t] .. to **limit** .. the the '**scope** [skop]....??..and '**maybe** [maɪbi] ..
be'cause they 'have [χæf] the '**main** [main] '**bulk** [bulk] of '**business** ['buzɪnəs]
(Rogerson-Revell, 2014, p. 147)

L2 pronunciation errors are rarely random attempts to produce unfamiliar sounds but typically reflect the learner's L1 phonological system, as a result of phonological transfer, often combined with orthographic transfer (see Chap. 2). It is relatively easy for L1 speakers to detect foreign accents, such as Italian- or Russian-accented English. Because L1 transfer plays an important role in L2 pronunciation, both at the segmental and suprasegmental levels (Ellis, 1994; Major, 2008; Odlin, 1989, chap. 6), knowledge of the learner's L1 phonological system can be a useful resource for teachers when diagnosing errors. However, as has long been known (Altenberg & Vago, 1983; Hecht & Mulford, 1987/1980; Richards, 1971; Tarone, 1978), many types of errors are not based on L1 transfer but rather on language universals and other natural learning processes, such as simplification and overgeneralization (which children may also apply in their language learning), as well as learners' incorrect hypotheses about the L2 based on their limited experience (see Chap. 2). In addition, the errors made by learners coming from the same L1 background show considerable variation, as reported, for example, by Altenberg and Vago (1983) for the L2 English speech of two adult L1 Hungarian speakers as elicited by a reading passage and by Munro and Derwing (2008) in a longitudinal study of instructed English vowel acquisition by speakers of Slavic languages. Although the learner's L1 plays a part in L2 phonological acquisition, so that contrast of the L1 and the L2 can predict some error types, the processes involved are diverse and complex, and sometimes similarities between languages predict pronunciation difficulties better than points of greatest contrast (Wode, 1977).

Error Severity

The severity of a pronunciation error in terms of how much it affects communication and understanding is a first criterion to consider. Extreme deviations, as in those segments marked ?? in the transcript above, are obviously of concern, though in some cases it may not be possible to discover what the learner was trying to say. Even if a listener can hear what an L2 speaker says, if individual phonemes are pronounced inaccurately, the listener may not be able to match what is heard to a sound sequence that can be recognized as an intelligible message made up of known words and grammatical structures. Those errors that make it impossible for a listener to extract an intelligible message can be considered to be the most severe, and so are a high priority for instruction. Errors which do not interfere with a listener's ability to understand can be considered less severe and so of less immediate importance in instruction.

Error Tolerance

One consideration which is not directly linked to intelligibility is listeners' **degree of tolerance**, that is, how well or poorly an error or L1-influenced feature of an L2 learner's pronunciation is tolerated by listeners. For example, some L1 French speakers of English may pronounce /θ/ as [s] in words like *thin* and *thick* and /ð/ as [z] in words like *this* and *that*. Such substitutions—which in this case represent simplifications for reasons of ease of articulation by speakers of a language that does not have interdental fricatives—rarely cause intelligibility problems, and English spoken with a French accent is considered quite favorably by many listeners. Yet, because words like *this* and *that* are so frequent, the [z] pronunciation of these words may be irritating to some listeners, as might the fact that *thin* and *thick* pronounced with [s] makes them homophonous with *sin* and *sick*, possibly producing negative connotations even when meaning is not at issue. Some listeners might be irritated by a particular pronunciation feature, including a prosodic one such as shifted word stress or overuse of a particular intonation pattern, even though it does not impede understanding. The degree of tolerance may vary significantly for different pronunciation features and languages (of both listeners and speakers), and for different

listener and speaker groups and individuals, and is often associated with linguistic stereotyping and “identity markers” (Zuengler, 1988).

Functional Load

One concept that is increasingly seen as important for deciding teaching priorities is **functional** (or **phonemic**) **load**, a concept associated with twentieth-century discussions in linguistics of the likelihood of sound change, on the assumption that “if a particular contrast is little used in the language, its elimination [i.e., by sound change, MCP/PRR] will do less harm than the elimination of a contrast with a high functional load” (Hoenigswald, 1960, p. 79). In relation to pronunciation, this refers to the amount of work phonemic contrasts do in lexis and in the phonological system overall to distinguish meaning, for example, /p/ and /b/ in minimal pairs such as *pin* and *bin*. It has been suggested that some phonemic contrasts do more work or have a higher functional load than others, either because the contrasts are more frequently occurring in words or because they are more significant. For example, the /p, b/ contrast occurs frequently in minimal pairs, both in initial position (e.g., *pan, ban*) and final position (e.g., *tap, tab*). The phonemic contrast is also significant because it distinguishes many minimal pairs of the same type or word class (nouns in this case) and because the contrast is “stable,” that is, not subject to significant regional variation (Brown, 1991; Catford, 1987). In comparison, the /ð, d/ pair distinguishes relatively few minimal pairs, rarely in the same word class, and the contrast is not distinctive in some regional or national accents (e.g., in the Scouse accent of Liverpool and in some creole-influenced accents of West Africa and the Caribbean; Wells, 1982a, p. 371, 1982b, pp. 565, 640). It would therefore be considered a low functional load contrast.

Considerations of functional load such as in the LFC and Amalgam English and International English proposals of Jenkins (2000) and Cruttenden (2008) may figure in decisions about teaching priorities. In addition, recent investigations of L2 learners’ error patterns suggest that high functional load errors might warrant attention in deciding teaching priorities. Munro and Derwing’s (2006) study of Cantonese learners of English concluded that high functional load errors were much more significant than low functional load errors in terms of comprehensibility, thus suggesting that they might logically

be a focus of instruction. On the other hand, Kang and Moran's (2014) research into functional load [FL] and pronunciation errors in oral assessment provides evidence that "[h]igh FL errors decrease significantly as proficiency levels increase.... It is evident that both vowel and consonant errors decrease with proficiency but drastic changes were shown with high FL errors" (p. 183). One possible implication of this research is that high functional load errors might not need special attention for advanced learners, for whom attention to low functional load errors may have benefits in terms of improving their pronunciation proficiency, since these are the kinds of errors that are low in communicative utility and cue value and so can benefit by explicit, form-focused instruction. Despite the potential usefulness of this concept for deciding teaching priorities, there has, to date, been little application of empirical research findings to teaching practices.

Fluency

Fluency is obviously a key consideration in pronunciation teaching and, although it can be operationalized in different ways, it is also often a key criterion in pronunciation assessment (Chap. 6). As we have seen in Chap. 1, there are many components of phonological fluency, generally related to how sequential, prosodic aspects of pronunciation are managed to maintain the easy flow of connected speech with minimal gaps and disruptions. Lack of temporal fluency, such as through inappropriate placement or overuse of pauses, can substantially disrupt the flow of speech and make it difficult for the listener to hear speech in meaningful chunks. On the other hand, occasional hesitation pauses are not likely to have a significantly negative effect on the listener, unless they are unusually long (several seconds). Temporal and phonological fluency can be considered key goals in language learning and teaching since they are key to listener comprehension and evaluation of a speaker. In addition, achieving a comfortable level of phonological fluency and automaticity makes it possible for a speaker to put pronunciation on "auto-pilot" (Chap. 1) and so focus on other aspects of communication. A focus on phonological fluency, which involves connected speech, also helps relate the teaching of pronunciation to the teaching of other aspects of spoken language, such as phrase structure, clause structure, and discourse structure, and how these are realized prosodically.

Impact

An aspect of pronunciation which is rarely prioritized in teaching or assessment is **impact**. This relates to the strategic or performance function of pronunciation, that is, the speaker's ability to use phonological features to make speech noteworthy to a listener or audience. As Rogerson-Revell (2011) explains, "A speaker may be intelligible to a listener without being very fluent. Similarly, a speaker may be intelligible and fluent without having much impact" (p. 246). There are many aspects of language which can be used to add impact to speech, including the use of metaphor or specific rhetorical devices (such as "two-part contrasts" and "three-part lists;" see discussion in Chap. 7 of such devices in Margaret Thatcher's speech). Proficient, articulate speakers also make careful use of prosodic features, including pauses, pitch range, and tone choice, to make a positive impression on an audience. How much priority is given to this aspect of pronunciation will ultimately depend on a learner's needs and proficiency level. Again, as Rogerson-Revell (2011) describes, "... someone who can produce words and phrases with reasonable accuracy and fluency but who does not make very much contrast between stressed and unstressed syllables may be competent enough to carry out a good variety of face-to-face interactions in English but not have enough clarity or impact to be a TV news announcer" (p. 247).

Teachability/Learnability

Another criterion which teachers do not always consider but which is worth bearing in mind is the **teachability–learnability** of a particular pronunciation feature, or the amount of effort involved, both for the learner and the teacher, in making an improvement in that feature. For instance, the effort made and the improvement achieved by teaching contrastive stress might be more worthwhile than attempting to teach the attitudinal function of tone choice. The teachability–learnability distinction is well established in relation to language teaching in general, as discussed by Allwright (1984) in "Why don't learners learn what teachers teach?" with reference to the observation that classroom teaching does not necessarily lead to classroom learning. A teachability–learnability scale has been proposed for pronunciation by Dalton and Seidlhofer (1994), who suggest:

Some things, say the distinction between fortis and lenis consonants, are fairly easy to describe and generalize—they are teachable. Other aspects, notably the attitudinal function of intonation, are extremely dependent on individual circumstances and therefore nearly impossible to isolate out for direct teaching. In other words, some aspects might better be left for learning without teacher intervention. (pp. 72–73)

While we agree with the general point that some things are easier to teach than others, it is not necessarily true that because a language feature is highly context-dependent it is for this reason “nearly impossible to isolate out for direct teaching,” given such methodologies as concordancing and digital manipulation of the speech signal. Moreover, provision of suitable context can make something highly memorable and salient, thus highly teachable even if complex. In addition, complex aspects of language and communication can have high teachability if learners have high interest and motivation to learn them.

Specific Learner Needs and Goals

Learners’ proficiency level and their overall goal for language learning will come to bear on any decisions about teaching priorities. Errors or problems in areas of language that are of particular importance for the learner’s social, academic, or professional needs will need to be prioritized. For instance, working on intonation patterns might be a high priority for learners aiming to work in customer service related jobs such as call center advisors, while correct pronunciation of multisyllabic medical terms will be crucial for those working as medical professionals. The need for pronunciation in specific fields and types of work suggests that pronunciation competencies should be specified and taught within ESP courses. Obviously, pronunciation errors and problems selected as teaching priorities should relate to criteria used for formative or diagnostic classroom assessment. Such criteria are considered in Chap. 6, including those of Pennington’s (1996, pp. 256–257) “Pedagogical classification of pronunciation errors and problems.”

Table 3.1 summarizes the key factors a teacher might consider when deciding pronunciation teaching priorities.

Table 3.1 Factors involved in deciding pronunciation teaching priorities (adapted from Rogerson-Revell, 2011, p. 247)

Factor	Reason	Priority for teaching	Example
Error severity	The error causes communication breakdown.	High	Lack of vowel lengthening before a final voiced consonant (e.g., <i>peace</i> vs. <i>peas</i>)
	The error does not affect intelligibility.	Low	Phoneme substitution without lexical confusion (e.g., [d] for [ð] in <i>this</i> and <i>that</i>) Shift in word stress (e.g., <i>'democracy</i> vs. <i>de'mocracy</i>)
Functional load	The feature is a frequently occurring contrast.	High	Lack of phoneme contrast (e.g., between pronunciation of /e/ in <i>bet</i> and /æ/ in <i>bat</i>)
	This feature or contrast occurs rarely or not in some varieties of English.	Low	Phonetic substitution (e.g., use of a trilled [r] or uvular [ʀ] instead of English post-alveolar approximant or retroflex /r/)
Fluency	The feature disrupts the flow of speech considerably.	High	Frequent pauses not at phrasal or information-unit boundaries
	The feature disrupts the flow of speech slightly.	Low	Disfluency (e.g., occasional hesitations for word search)
Impact	The use/lack of use of this feature reduces the impact of the message considerably.	High	Lack of use of stress to signal sentence-level meaning differences (e.g., <i>it's been a good day</i> ; <i>let's go</i> vs. <i>It's been a good day</i> ; <i>let's go</i>) Insufficient distinction (or no) differentiation between stressed/unstressed words (e.g., <i>'when 'did 'he 'go 'to 'the 'bank</i>)

(continued)

Table 3.1 (continued)

Factor	Reason	Priority for teaching	Example
	The use/lack of use of this feature reduces the impact of the message slightly.	Low	Insufficient use or overuse of controlled pauses (e.g., overly rapid or slow delivery)
Degree of tolerance	The error is a source of irritation for the listener and requires a high degree of tolerance to accept or ignore.	High	Nonstandard tone choice (e.g., frequent use of final rising tone on statements) Incorrect word stress (e.g., 'develop instead of de'velop)
	Listeners are used to, or accept, this feature of foreign or regional accent.	Low	Lack of contrasting use of light /l/ (e.g., in <i>lip</i>) and dark /l/ [ɫ] (e.g., in <i>eel</i>)
Teachability–learnability	The effort involved merits the result achieved.	High	Stress placement (e.g., <i>it's my bag</i> vs. <i>it's my <u>bag</u></i>) Chunking into tone units or phrasal divisions (e.g., // <i>when you're ready</i> // <i>let me know</i> //)
	The effort invested is greater than the return on investment.	Low	Attitudinal functions of tone choice (e.g., of a rise-fall pattern vs. a simple fall: ^ <i>great</i> vs. ` <i>great</i>)
Goal	The feature would prevent reaching the level of competence required.	High	Lack of lengthening of vowels before final voiced consonants, for language teachers (e.g., <i>bed</i> vs. <i>bet</i>)
	The feature will not interfere with achieving the target.	Low	Lack of vowel reduction and nonuse of schwa for speakers in EIL contexts (e.g., <i>banana</i> [ba'næna])

Case Studies of Pronunciation Teaching Geared to Student Needs and Contexts of Language Use

In this section, we offer two examples of how pronunciation instruction might be designed to address the specific needs of students in the international contexts in which English is increasingly being used. The first case illustrates how a teacher might address the needs of students in an “expanding circle” country (Singapore) where a local variety of English is in widespread use. The second case study illustrates how the needs of L2 speakers who must communicate in an L1 environment, international medical graduates working in a U.S. hospital, might be addressed. These two studies provide a good contrast of pronunciation concerns and needs in these L2 and L1 contexts while also providing contrasts of a focus on prosody (Case Study 1) versus segmentals (Case Study 2). They also suggest ways in which research can inform practice, which is an important theme of this book.

Case Study 1: Pedagogy Focused on Prosody for L2 Students in an Expanding Circle Country (Low, 2006)

Our first illustrative case study is one singled out as a “model study” by Pennington (2015a) of pronunciation research with useful implications for practice. Eva Ling Low carried out research on different varieties of English in comparative perspective (Low, 2006), as a basis for making recommendations for teaching practice that would be geared to the needs of students with respect to those two varieties. The research was carried out in Singapore, where a New English variety of Singaporean English is well-established that has different conventions from those of GB or Southern British English for foregrounding and backgrounding information by use of prosody, as in the given/new distinction. The study demonstrated careful research design and use of modern technology; “a strong command of contemporary phonology, SLA theory and research, relevant classical and recent published literature, and the lan-

guages under study”; and “originality and value for the field” (Pennington, 2015a, p. 159), with research results specifically related to teaching recommendations.

Low (2006) found that the Singapore English speakers, in contrast to a sample of speakers of British English, produced given information with a much lesser lowering of pitch for given information and also with a different intonation contour overall. Low discusses the likelihood that the meaning or pragmatic force of a Singaporean’s utterance might be misunderstood by non-Singaporeans (specifically, native speakers of British English but potentially others as well) based on the lack of deaccenting or destressing of given information. It would therefore seem worthwhile to teach the use of prosody to background given information as contrastive with prosodic foregrounding of new information.

The question naturally arises as to how the Singaporean speakers signal the given/new distinction other than through prosody. Low notes that final pragmatic markers (which are common in Chinese) often perform this function. The final marker *lah* marks information as new (Example 1), whereas the final marker *what* marks information as given, or something that should already be known (Example 2):

(1) *It’s the man, lah, I tell you it’s him.*

(2) A: *Why do you like the piano?*

B: *I like, what?*

Low’s (2006) research had a pedagogical purpose and incorporates recommendations for teaching that reflect a modern view of pronunciation involving student identity, agency, and the choices L2 speakers can make in interacting with speakers from different groups. A syllabus for pronunciation, in her view,

...should include aspects of nuclear and contrastive stress placement, but also ...should highlight features that differ from native varieties as differences not errors. At the same time, how the student’s own variety

of English conveys the same discursial intent should also be highlighted and not ignored. An effective way [to] highlight discursial intent would be to design conversations that would cue the presence of deaccenting in native varieties of English and to ask speakers of nonnative varieties to produce these conversations. The teacher can also encourage the pupils to paraphrase the conversations with their own words as long as they keep the content exactly the same. The teacher can then observe how students make distinctions between new and given information and, if possible, use this information for classroom instruction. (pp. 757–758)

Thus, Low recommends a focus on form which recognizes that the students command more than one variety of English and explicitly builds their awareness of varietal differences and how they can impact communication.

This study illustrates how investigations of linguistic differences can provide a basis for instruction which both values and explicitly addresses the features of students' L2 English variety in relation to an L1 variety. It thus offers direction for designing instruction based on the reality of English as an international language that has created local varieties which play a major role in students' communication and so cannot rightfully be ignored in teaching and learning. In such cases, it can be of value to explicitly compare and contrast the different models on which pronunciation can be based, to raise students' awareness and appreciation of their own ingroup variety at the same time as they learn about the native speaker model, thus improving their command of the latter while also enhancing their agency and ability to use different varieties based on purpose and circumstances. Conducting systematic research to discover the differing models and patterns of specific aspects of pronunciation, such as for marking given/new contrasts as done by Low (2006), is a valuable experience for language teachers and trainers as well, in sensitizing them to the communication difficulties and hurdles their students face and to the functional value of different varieties. Such research can then lead into action research or other instructional interventions to bring that new sensitivity and understanding into their teaching or training efforts.

Case Study 2: Pedagogy Focused on Segmental Phonology for International Medical Graduates Practicing in an L1 Context (Labov & Hanau, 2011)

Work by Joanna Labov, a pronunciation specialist, and Cheryl Hanau, a medical administrator and professor of pathology and laboratory medicine, demonstrates the value of a pronunciation focus for **international medical graduates** (IMGs) in a course specifically geared to address their work-related pronunciation problems in an L1 English environment (Labov & Hanau, 2011). The study serves as a model of a collaborative project focused on ESP pronunciation training and research, and of the course process and outcomes, that may give ideas for pronunciation R&D (research and development) in other work-related settings (see Chap. 7).

The specific learners on the course were doctors working in a pathology residency at an urban hospital who were required by their departmental administrator (Hanau) to attend the course “because their pronunciation was often incomprehensible to the medical secretaries who transcribed their dictations” (p. 262)—a crucial ability, given that a main activity of pathologists is to dictate reports describing the appearance of specimens removed during surgery and evaluated by them on that basis as healthy or not. These dictated reports are, according to Labov and Hanau (2011, p. 262), “critical for the health of the patient”: they need to be accurate and comprehensible since they represent, in effect, “verbal photographs” that other doctors will use as a basis for “medically important decisions.”

The seven pathology residents required to attend the course, five Chinese and two Peruvians who were all fluent English speakers but had speech which the researchers described as “fossilized” or “stabilized” (p. 267), were highly motivated to improve their pronunciation. The course was designed within a research study whose goal was “to determine whether doctors who have clear evidence of stabilized phonology can improve their English pronunciation through intensive study” (Labov & Hanau, 2011, p. 269). In addition, data from the research study was intended to provide information on the frequency and types of errors in pronunciation which the doctors experienced that could feed back into training.

An intensive 10-week course was designed with the goal of “improv[ing] the comprehensibility of the doctors’ dictations of their specimen reports by achieving greater clarity and accuracy of their speech...as measured by the medical secretaries’ comprehension of the dictations” (p. 266). The needs-focused, ESP orientation of the course meant that it was “task-based, communicative, and student-centered” (Labov & Hanau, 2011, p. 267), with close alignment to the learners’ specific work-related concerns of oral dictation of pathology reports. Course sessions involved either group lessons or individual tutorials of 90 minutes once a week. Each doctor supplied an audio recording and a transcription of a dictated pathology report on the first day of class, and they were required to record themselves reading the same transcription in weeks 6 and 9 of the course. The tapes made in weeks 1 and 9 would then be used to assess the effectiveness of the course.

The course began with diagnostics and goal-setting in week 1 and included an overview in week 4 of common issues in communication having to do with segmentals (vowel height, duration, and contrasts; place and manner of articulation; voicing; and consonant contrasts) and more issues in communication having to do with suprasegmentals (intonation, stress, and other aspects of prosody, in addition to pausing and rate of speech) in week 7. Individual tutorials were given in weeks 2, 5, and 8 to address each doctor’s needs preparatory to an in-class dictation presentation in each of the weeks that followed (i.e., weeks 3, 6, and 9). These dictation presentations were then given to the medical secretaries to transcribe. Week 10 included individual conferences and wrap-up of the course.

The course text was Orion’s (1999) *Pronouncing American English: Sounds, Stress, and Intonation*, which uses physical displays of articulatory contrasts in place and manner and instructions for articulation of each phoneme that were considered especially appropriate for doctors. These graphics and instructions were supplemented by tactile activities that allowed them to feel tongue positions and movements and visual demonstrations such as using “tissues to teach syllable-initial aspiration, mirrors to view the oral cavity, and rubber bands to demonstrate vowel duration preceding voiced and voiceless consonants” (Labov & Hanau, 2011, p. 268). The relevance of the course to the doctors’ work was reinforced by using their own dictations as part of the course material and selecting specific words from these to work on in the whole-group sessions as well

as in the individual tutorials. In addition, the symbols of the **International Phonetic Alphabet** (IPA), which were already known by the Chinese doctors and were taught to the two Peruvian doctors, were used to help the learners pronounce consonants and vowels accurately.

The effectiveness of instruction was evaluated based on the week 1 and week 9 tapes of four of the Chinese doctors and one of the Peruvian doctors, since there were problems with the recordings of the two other doctors. The tapes were randomized and given to four experienced L1 English medical secretaries to transcribe. 72 (2%) of the words produced in the week 1 tapes were unintelligible to the secretaries, and 52 (1.4%) were misunderstood by them. This is considered to be a high error rate for medical transcription that is mainly due to the residents' pronunciation. Of interest is the fact that only 25% of the unintelligible words and 19% of the misunderstood words were technical words, such as *infarction* heard as *infection*; the remainder were non-technical words such as *edge*, which was judged unintelligible, and *yellow*, which was misheard as *gelatin*. There was improvement in comprehensibility of words from week 1 to week 9 that was statistically significant for words that were explicitly taught and overall, including those which were not explicitly taught, thus indicating that the course had been effective in its goal of improving the doctor's work-related pronunciation.

The research findings of this teaching intervention stands as evidence that significant improvement in pronunciation can be achieved in a relatively limited time under conditions that include:

- a combination of in-class and individualized teaching based on learner needs and work requirements;
- instructional activities and materials that include explicit form-focused explanations and articulatory instructions, graphics, and demonstrations, together with specific work-related communicative tasks;
- a pronunciation specialist teacher; and
- highly motivated learners.

This study provides direction for other small-scale, and potentially larger scale, curriculum development and research projects involving targeted pronunciation interventions for specific populations of L2 speakers based on analysis of their work-related communication problems.

Concluding Remarks

In this chapter, we have reviewed the historical background to pronunciation teaching related to changing views of language and of learning within the larger theoretical and practical concerns of linguistics and applied linguistics, as influenced by education and psychology, that have initiated and supported various orientations to language teaching within second language acquisition, bilingual education, and multilingualism. We have also considered some of the key issues involved in planning and implementing pronunciation teaching, including selecting appropriate models and goals and identifying teaching priorities. Awareness of developments in pronunciation teaching and some of the theoretical and practical issues that have arisen can help situate pronunciation teaching not only within the broader curriculum but also in terms of the needs and concerns of individual learners. Two case studies have served to illustrate how features of both the international and local contexts define English usage and form a basis for decisions and curriculum design for pronunciation pedagogy.

Evidence is increasingly showing that the context of learning should directly impact how pronunciation is addressed, and decisions about models, goals, and teaching priorities should reflect this. Differences in learner needs for English and the predominant use of English for international communication mean that it is not realistic to take a monolithic approach to pronunciation teaching. Assumptions about the supremacy of native speaker models and targets, particularly in ELF contexts, have rightly been questioned and have consequently influenced how teaching and learning priorities are established. Users of English who interact professionally with native speakers may need to adjust to an L1 English or “neutral” pronunciation model, and may also need to develop advanced competence in aspects of pronunciation that go considerably beyond intelligibility to affect interactive dynamics and communicative effectiveness; but for users who communicate predominantly and in limited ways with other L2 English speakers, such a model may not be appropriate or necessary (Jenkins, 2000). Given the differing needs for an L2 and the differing types of communication and communicative contexts in which learners will participate in their lives, issues remain regarding what an

appropriate model for pronunciation teaching is and what features of that model are to be emphasized in instruction.

Curriculum reflecting important matters of language in everyday life should not be dictated based on limited knowledge or outmoded ideas and misconceptions. As Levis (2005) concludes, “findings indicate that teaching pronunciation is only partially a pedagogical decision, and that old assumptions are ill-suited to a new reality” (p. 376). Decisions surrounding the teaching of pronunciation need to be based on informed knowledge and to be taken thoughtfully, in the light of the new reality of the rapidly evolving international contexts in which English is being used and in consideration of the factors, practical as well as theoretical, impacting curriculum decisions that have been reviewed in this chapter and will be augmented and increasingly contextualized in the chapters to come, and then revisited in the final chapter.

Notes

1. It should be noted that the model for HVPT in the original Logan et al. (1991) study and in many other studies implementing HVPT since the time of that study offers training for L2 learners using stimuli from a variety of speakers and phonetic environments, but not specifically varying such features as accent or linguistic variety, genre, or type of discourse.
2. Cruttenden uses the term “English as an additional language” to refer to any use of language other than as an L1 (Cruttenden, 2008, 2014).

References

- Allwright, R. L. (1984). Why don't learners learn what teachers teach? The interaction hypothesis. In D. M. Singleton & D. G. Little (Eds.), *Language learning in formal and informal contexts* (pp. 3–18). Dublin: IRAAL.
- Altenberg, E. P., & Vago, R. M. (1983). Theoretical implications of an error analysis of second language phonology production. *Language Learning*, 33(4), 427–448. <https://doi.org/10.1111/j.1467-1770.1983.tb00943.x>
- Beinhoff, B. (2013). *Perceiving identity through accent: Attitudes towards non-native speakers and their accents in English*. Oxford and Bern: Peter Lang.

- Block, D. (2003). *The social turn in second language acquisition*. Edinburgh: Edinburgh University Press.
- Bradac, J. J. (1990). Language attitudes and impression formation. In H. Giles & W. P. Robinson (Eds.), *Handbook of language and social psychology* (pp. 387–412). Chichester, UK: Wiley.
- Brown, A. (1991). *Pronunciation models*. Singapore: Singapore University Press.
- Brumfit, C. J., & Johnson, K. (1979). *The communicative approach to language teaching*. Oxford: Oxford University Press.
- Cargile, A. C., Giles, H., Ryan, E. B., & Bradac, J. J. (1994). Language attitudes as a social process: A conceptual model and new directions. *Language & Communication*, 14(3), 211–236. [https://doi.org/10.1016/0271-5309\(94\)90001-9](https://doi.org/10.1016/0271-5309(94)90001-9)
- Carrie, E. (2017). British is professional, American is urban: Attitudes towards English reference accents in Spain. *International Journal of Applied Linguistics*, 27(2), 427–447. <https://doi.org/10.1111/ijal.12139>
- Catford, J. C. (1987). Phonetics and the teaching of pronunciation: A systematic description of English phonology. In J. Morley (Ed.), *Current trends in pronunciation: Practices anchored in theory* (pp. 87–100). Washington, DC: TESOL.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Chun, D. (2007). Technological advances in researching and teaching phonology. In M. C. Pennington (Ed.), *Phonology in context* (pp. 274–229). Basingstoke, UK and New York: Palgrave Macmillan.
- Collins, B., & Mees, I. M. (2013). *Practical phonetics and phonology: A resource book for students* (3rd ed.). Abingdon, UK and New York: Routledge.
- Coste, D., Moore, D., & Zarate, G. (2009/1997). *Plurilingual and pluricultural competence*. Strasbourg: Council of Europe, Language Policy Division. Studies towards a Common European Framework of Reference for language learning and teaching. Revised from original version that appeared in French, 1997. Retrieved January 3, 2017, from https://www.coe.int/t/dg4/linguistic/Source/SourcePublications/CompetencePlurilingue09web_en.pdf
- Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Strasbourg: Council of Europe, Language Policy Division. Retrieved January 3, 2017, from <http://www.coe.int/lang>. Also published in its English language version under the same title by Cambridge University Press.
- Cruttenden, A. (2008). *Gimson's pronunciation of English* (7th ed.). London: Hodder Education.
- Cruttenden, A. (2014). *Gimson's pronunciation of English* (8th ed.). London: Hodder Education.

- Dalton, C., & Seidlhofer, B. (1994). *Pronunciation*. Oxford: Oxford University Press.
- Dalton-Puffer, C. (2007). *Discourse in content and language integrated learning (CLIL) classrooms*. Amsterdam and Philadelphia: John Benjamins.
- Dauer, R. M. (2005). The Lingua Franca Core: A new model for pronunciation instruction? *TESOL Quarterly*, 39(3), 543–550. <https://doi.org/10.2307/3588494>
- Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. Amsterdam and Philadelphia: John Benjamins.
- Derwing, T. M., & Rossiter, M. J. (2002). ESL learners' perceptions of their pronunciation needs and strategies. *System*, 30(2), 155–166. [https://doi.org/10.1016/S0346-251X\(02\)00012-X](https://doi.org/10.1016/S0346-251X(02)00012-X)
- Dudley-Evans, T., & St. John, M. J. (1998). *Developments in English for specific purposes: A multi-disciplinary approach*. Cambridge, UK: Cambridge University Press.
- Eckert, P. (2000). *Linguistic variation as social practice*. Oxford: Blackwell.
- Eckert, P., & Rickford, J. R. (2001). *Style and sociolinguistic variation*. Cambridge: Cambridge University Press.
- Edwards, J. (1994). *Multilingualism*. London: Penguin Books.
- Ellis, N. C. (1994). Implicit and explicit processes in language acquisition: An introduction. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 1–32). London: Academic Press.
- Ellis, R. (2003). *Task-based language teaching and learning*. Oxford: Oxford University Press.
- English Language Services. (1964). *English 900*. New York: Collier Macmillan.
- Esling, J. H. (1987). Methodology for voice setting awareness in language classes. *Revue de Phonétique Appliquée*, 85, 449–473.
- Flege, J. E., Munro, M., & MacKay, I. R. A. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America*, 97(5), 3125–3134. <https://doi.org/10.1121/1.413041>
- Fought, C. (2006). *Language and ethnicity: Key topics in sociolinguistics*. New York: Cambridge University Press.
- García, O. (2009). *Bilingual education in the 21st century: A global perspective*. Oxford: Wiley Blackwell.
- Genesee, F. (1987). *Learning through two languages: Studies of immersion and bilingual education*. Cambridge, MA: Newbury House.
- Gilbert, J. B. (2012). *Clear speech: Pronunciation and listening comprehension in North American English* (4th ed.). New York: Cambridge University Press.

- Giles, H. (1970). Evaluative reactions to accents. *Educational Review*, 22(3), 211–227. <https://doi.org/10.1080/0013191700220301>
- Giles, H., Bourhis, R. Y., & Taylor, D. M. (1977). Towards a theory of language in ethnic group relations. In H. Giles (Ed.), *Language, ethnicity and inter-group relations* (pp. 307–348). London: Academic Press.
- Giles, H., & Sassoon, C. (1983). The effect of speaker's accent, social class background and message style on British listeners' social judgments. *Language and Communication*, 3(3), 305–313. [https://doi.org/10.1016/0271-5309\(83\)90006-X](https://doi.org/10.1016/0271-5309(83)90006-X)
- Gluszek, A., & Dovidio, J. F. (2010). The way they speak: A social psychological perspective on the stigma of nonnative accents in communication. *Personality and Social Psychology Review*, 14(2), 214–237. <https://doi.org/10.1177/1088868309359288>
- Graddol, D. (2006). *English next: Why global English may mean the end of "English as a Foreign Language"*. The British Council. Retrieved September 11, 2017, from <https://englishagenda.britishcouncil.org/continuing-professional-development/cpd-researchers/english-next>
- Halliday, M. A. K. (1970). Language structure and language function. In J. Lyons (Ed.), *New horizons in linguistics* (pp. 140–165). Harmondsworth, UK: Penguin.
- Halliday, M. A. K. (1973). *Explorations in the functions of language*. London: Edward Arnold.
- Hammerly, H. (1982). *Synthesis in second language teaching: An introduction to linguistics [sic]*. Blaine, WA: Second Language Publications.
- Harris, Z. (1951). *Methods in structural linguistics*. Chicago: University of Chicago Press.
- Hecht, B. F., & Mulford, R. (1987/1980). The acquisition of a second language phonology: Interaction of transfer and developmental factors. In G. Ioup & S. Weinberger (Eds.), *Interlanguage phonology: The acquisition of a second language sound system* (pp. 213–228). New York: Harper & Row. Originally published 1980 in *Papers and Reports in Child Language Development*, 18, 61–74.
- Hockett, C. (1958). *A course in modern linguistics*. New York: Macmillan.
- Hoenigswald, H. (1960). *Language change and linguistic reconstruction*. Chicago and London: University of Chicago Press.
- Honey, J. (1989). *Does accent matter? The Pygmalion factor*. London: Faber and Faber.
- Hornby, A. S. (1950). The situational approach to language teaching: A series of three articles in English. *Language Teaching*, 4, 98–104, 121–128, 150–156.

- Hymes, D. (1972). On communicative competence. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics* (pp. 269–293). Harmondsworth, UK: Penguin.
- Isaacs, T., & Trofimovich, P. (2012). Deconstructing comprehensibility. *Studies in Second Language Acquisition*, 34(4), 475–505. <https://doi.org/10.1017/S0272263112000150>
- 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. <https://doi.org/10.1093/applin/23.1.83>
- Jenkins, J. (2005). Implementing an international approach to English pronunciation: The role of teacher attitudes and identity. *TESOL Quarterly*, 39(3), 535–543. <https://doi.org/10.2307/3588493>
- Jenkins, J. (2012). English as a lingua franca from the classroom to the classroom. *English Language Teaching Journal*, 66(4), 486–494. <https://doi.org/10.1093/elt/ccs040>
- Jenner, B. (1989). Teaching pronunciation: The common core. *Speak Out!* 4, 2–4. Kent, UK: IATEFL Pronunciation Special Interest Group.
- Johnstone, B. (2009). Stance, style and the linguistic individual. In A. Jaffe (Ed.), *Sociolinguistic perspectives on stance* (pp. 29–52). New York: Oxford University Press.
- Kachru, B. B. (1985). Standards, codification and sociolinguistic realism: The English language in the outer circle. In R. Quirk & H. Widdowson (Eds.), *English in the world: Teaching and learning the language and literatures* (pp. 11–30). Cambridge: Cambridge University Press.
- Kang, O., & Moran, M. (2014). Functional loads of pronunciation features in nonnative speakers' oral assessment. *TESOL Quarterly*, 48(1), 176–187. <https://doi.org/10.1002/tesq.152>
- Kenworthy, J. (1989). *Teaching English pronunciation*. London: Longman.
- Kibishi, H., Hirabayashi, K., & Nakagawa, S. (2014). A statistical method of evaluating the pronunciation proficiency/intelligibility of English presentations by Japanese speakers. *ReCALL*, 27(1), 58–83. <https://doi.org/10.1017/S0958344014000251>
- Krashen, S. (1982). *Principles and practice in second language acquisition*. Oxford: Pergamon Press.
- Krashen, S. (1985). *The input hypothesis: Issue and implications*. New York: Longman.
- Krashen, S., & Terrell, T. (1983). *The natural approach: Language acquisition in the classroom*. Oxford: Pergamon.

- Kuo, I.-C. (2006). Addressing the issue of teaching English as a lingua franca. *ELT Journal*, 60(3), 213–221. <https://doi.org/10.1093/elt/ccl001>
- Labov, J., & Hanau, C. (2011). Pronunciation as life and death: Improving the communication skills of non-native English speaking pathologists. In B. J. Hoekje & S. M. Tipton (Eds.), *English language and the medical profession: Instructing and assessing the communication skills of international physicians* (pp. 261–285). Bingley, UK; Leiden and Boston: Emerald; Brill.
- Labov, W. (1966). *The social stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.
- Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- Lado, R. (1964). *Language teaching: A scientific approach*. New York: McGraw-Hill.
- Levis, J. M. (2005). Changing contexts and shifting paradigms. *TESOL Quarterly*, 39(3), 369–377. <https://doi.org/10.2307/3588485>
- Levis, J. M. (2016). The interaction of research and pedagogy (Editorial). *Journal of Second Language Pronunciation*, 2(1), 1–7. <https://doi.org/10.1075/jslp.2.1.001lev>
- Levis, J. M., & Barriuso, T. A. (2012). Nonnative speakers' pronunciation errors in spoken and read English. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 3rd pronunciation in second language learning and teaching conference*, September 2011 (pp. 187–194). Ames, IA: Iowa State University.
- Levis, J. M., & Grant, L. (2003). Integrating pronunciation into ESL/EFL classrooms. *TESOL Journal*, 12(2), 13–19. <https://doi.org/10.1002/j.1949-3533.2003.tb00125.x>
- Li Wei. (2011). Moment analysis and translanguaging space. *Journal of Pragmatics*, 43(5), 1222–1235. <https://doi.org/10.1016/j.pragma.2010.07.035>
- Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /ɹ/ and /l/: A first report. *Journal of the Acoustical Society of America*, 89(2), 874–886. <https://doi.org/10.1121/1.1894649>
- Long, M. (2015). *Second language acquisition and task-based language teaching*. Malden, MA and Oxford, UK: Wiley Blackwell.
- 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. <https://doi.org/10.2307/40264306>
- Major, R. C. (2008). Transfer in second language phonology: An overview. In J. G. Hansen Edwards & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 63–94). Amsterdam and Philadelphia: John Benjamins.

- Mey, J. (1981). Right or wrong, my native speaker. In F. Coulmas (Ed.), *A festschrift for native speaker* (pp. 69–84). The Hague: Mouton.
- Mompean, J. A. (2004). Options and criteria for the choice of an English pronunciation model in Spain. In J. Anderson, J. M. Oro Cabanas, & J. Varela Zapata (Eds.), *Linguistic perspectives from the classroom: Language teaching in a multicultural Europe* (pp. 1043–1059). Santiago de Compostela, Spain: Servizo de Publicacións e Intercambio Científico da Universidade de Santiago de Compostela Retrieved April 12, 2018, from <https://www.researchgate.net/publication/259820691>
- Mompean, J. A. (2008). Consumers' preferences and the choice of English pronunciation models. In R. Monroy & A. Sánchez (Eds.), *25 years of applied linguistics in Spain: Milestones and challenges* (pp. 959–964). Murcia, Spain: Editum.
- Moussu, L., & Llurda, E. (2008). Non-native English-speaking English language teachers: History and research. *Language Teaching*, 41(3), 315–348. <https://doi.org/10.1017/S0261444808005028>
- Munro, M. J. (2003). A primer on accent discrimination in the Canadian context. *TESL Canada Journal*, 20(2), 38–51. <https://doi.org/10.18806/tesl.v20i2.947>
- Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 45(1), 73–97. <https://doi.org/10.1111/j.1467-1770.1995.tb00963.x>
- Munro, M. J., & Derwing, T. M. (2006). The functional load principle in ESL pronunciation instruction: An exploratory study. *System*, 34(4), 520–531. <https://doi.org/10.1016/j.system.2006.09.004>
- Munro, M. J., & Derwing, T. M. (2008). Segmental acquisition in adult ESL learners: A longitudinal study of vowel production. *Language Learning*, 58(3), 479–503. <https://doi.org/10.1111/j.1467-9922.2008.00448.x>
- Munro, M. J., Derwing, T. M., & Morton, S. L. (2006). The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition*, 28(1), 111–131. <https://doi.org/10.1017/S0272263106060049>
- Odlin, T. (1989). *Language transfer: Cross-language influence in language learning*. Cambridge: Cambridge University Press.
- Orion, G. F. (1999). *Pronouncing American English: Sounds, stress, and intonation*. Boston: Heinle & Heinle.
- Ortega, L. (2013). SLA for the 21st century: Disciplinary progress, transdisciplinary relevance, and the bi/multilingual turn. *Language Learning*, 63(1), 1–24. <https://doi.org/10.1111/j.1467-9922.2012.00735.x>

- Otsuji, E., & Pennycook, A. (2010). Metrolingualism: Fixity, fluidity and language in flux. *International Journal of Multilingualism*, 7(3), 240–254. <https://doi.org/10.1080/14790710903414331>
- Pennington, M. C. (1989). Applications of computers in the development of speaking and listening proficiency. In M. C. Pennington (Ed.), *Teaching languages with computers: The state of the art* (pp. 97–121). La Jolla, CA: Athelstan.
- Pennington, M. C. (1996). *Phonology in English language teaching: An international approach*. London and New York: Longman.
- Pennington, M. C. (2015a). Research, theory, and practice in second language phonology: A review and directions for the future. In J. A. Mompean & J. Fouz-González (Eds.), *Investigating English pronunciation: Trends and directions* (pp. 149–173). Basingstoke, UK and New York: Palgrave Macmillan.
- Pennington, M. C. (2015b). From L2 phonology to pluriphonology: A new perspective on pronunciation theory, research, and practice. Presentation at the *American Association for Applied Linguistics conference*, Toronto, March 2015.
- Pennington, M. C. (2018). Identity in language learning. In J. C. Richards & A. Burns (Eds.), *Cambridge guide on second language learning* (pp. 91–98). New York: Cambridge University Press.
- Pennington, M. C., & Esling, J. H. (1996). Computer-assisted development of spoken language skills. In M. C. Pennington (Ed.), *The power of CALL* (pp. 153–189). Houston: Athelstan.
- Pennington, M. C., & Richards, J. C. (2016). Teacher identity in language teaching: Integrating personal, contextual, and professional factors. *RELC Journal*, 47(1), 1–19.
- Pica, T. (1984). Pronunciation activities with an accent on communication. *English Teaching Forum*, 22(3), 2–6. Retrieved from <https://babel.hathitrust.org/cgi/pt?id=nncl.cu14472120>
- Porter, D., & Garvin, S. (1989). Attitudes to pronunciation in EFL. *Speak Out!*, 5, 8–15. Kent, UK: IATEFL Pronunciation Special Interest Group.
- Rampton, B. (1995). *Crossing: Language and ethnicity among adolescents*. London: Longman.
- Ramsey, P. J. (2012). *The bilingual school in the United States: A documentary history*. Charlotte, NC: New Age Publishing.
- Richards, J. C. (1971). A non-contrastive approach to error analysis. *English Language Teaching*, 25(3), 204–219. <https://doi.org/10.1093/elt/XXV.3.204>
- Richards, J. C., & Rodgers, T. S. (2014). *Approaches and methods in language teaching* (3rd ed.). New York: Cambridge University Press.
- Rogerson, P., & Gilbert, J. B. (1990). *Speaking clearly*. Cambridge: Cambridge University Press.

- Rogerson-Revell, P. (2011). *English phonology and pronunciation teaching*. London: Bloomsbury.
- Rogerson-Revell, P. (2014). Pronunciation matters: Using English for international business communication. In R. van den Doel & L. Rupp (Eds.), *Pronunciation matters* (pp. 137–156). Amsterdam: VU Uitgeverij.
- Savignon, S. (1983). *Communicative competence; theory and classroom practice*. Reading, MA: Addison-Wesley.
- Scales, J., Wennerstrom, A., Richards, D., & Wu, S. (2006). Language learners' perceptions of accent. *TESOL Quarterly*, 40(4), 715–738. <https://doi.org/10.2307/40264305>
- Seidlhofer, B. (2011). *Understanding English as a lingua franca*. Oxford: Oxford University Press.
- Sewell, A. (2009). World Englishes, English as a *lingua franca*, and the case of Hong Kong English. *English Today*, 25(1), 37–43. <https://doi.org/10.1017/S0266078409000066>
- Sewell, A. (2013). English as a lingua franca: Ontology and ideology. *ELT Journal*, 67(1), 3–10. <https://doi.org/10.1093/elt/ccs061>
- Shaw, P. (2008). Spelling, accent and identity in computer-mediated communication. *English Today*, 24(2), 42–49. <https://doi.org/10.1017/S0266078408000199>
- Skinner, B. F. (1957). *Verbal behavior*. New York: Appleton-Century-Crofts.
- Smith, L., & Nelson, C. (1985). International intelligibility of English: Directions and resources. *World Englishes*, 4(3), 333–342. <https://doi.org/10.1111/j.1467-971X.1985.tb00423.x>
- Sung, C. C. M. (2013). English as a lingua franca and English language teaching: A way forward. *ELT Journal*, 67(3), 350–353. <https://doi.org/10.1093/elt/cct015>
- Tarone, E. E. (1978). The phonology of interlanguage. In J. C. Richards (Ed.), *Understanding second and foreign language learning* (pp. 15–33). Rowley, MA: Newbury House.
- Terrell, T. D. (1977). A natural approach to second language acquisition and learning. *Modern Language Journal*, 61(7), 325–336. <https://doi.org/10.2307/324551>
- Terrell, T. D. (1982). The natural approach to language teaching: An update. *Modern Language Journal*, 66(2), 121–132. <https://doi.org/10.2307/326380>
- Timmis, I. (2002). Native speaker norms and international English: A classroom view. *ELT Journal*, 56(3), 240–249. <https://doi.org/10.1093/elt/56.3.240>
- Trudgill, P. (1974). *The social differentiation of English in Norwich*. Cambridge: Cambridge University Press.

- Ur, P. (2012). English as an international language: Implications for classroom teaching and teaching materials. Lecture at Bar-Ilan University. Retrieved from <https://www.youtube.com/watch?v=yTidAm0dRR0>
- Walker, R. (2011). *Teaching the pronunciation of English as a lingua franca*. Oxford: Oxford University Press.
- Wells, J. C. (1982a). *Accents of English 2: The British Isles*. Cambridge: Cambridge University Press.
- Wells, J. C. (1982b). *Accents of English 3: Beyond the British Isles*. Cambridge: Cambridge University Press.
- Wells, J. C. (1997). Whatever happened to received pronunciation? In C. Medina Casado & C. Soto Palomo (Eds.), *II Jornadas de Estudios Ingleses* (pp. 19–28). Spain: Universidad de Jaén. Retrieved February 26, 2018, from <http://www.phon.ucl.ac.uk/home/wells/pubs.htm>
- Widdowson, H. G. (1978). *Teaching language as communication*. Oxford: Oxford University Press.
- Widdowson, H. G. (1994). The ownership of English. *TESOL Quarterly*, 28(2), 377–389. <https://doi.org/10.2307/3587438>
- Wode, H. (1977). The L2 acquisition of /r/. *Phonetica*, 34(3), 200–217. <https://doi.org/10.1159/000259877>
- Zuengler, J. (1988). Identity markers and L2 pronunciation. *Studies in Second Language Acquisition*, 10(1), 33–49. <https://doi.org/10.1017/S027226310000694X>



4

Pronunciation in the Classroom: Teachers and Teaching Methods

Introduction

The previous chapter considered some issues relating to what to teach, but that leaves the questions of when and how to teach pronunciation as yet unaddressed. As we have stated before, pronunciation still remains relatively marginalized in the general language curriculum and often tends to be taught in a rather ad-hoc way, with little pre-planning and with the teacher focusing primarily on providing feedback on errors (Foote, Trofimovich, Collins, & Urzúa, 2013).

Many teachers and researchers are aware that much less pronunciation teaching goes on in classrooms than optimal and believe that pronunciation “gets too small a piece of the language teaching pie” (Levis, 2007, p. 197). This situation is not surprising given the many demands on most language teachers’ time and the pressures often presented by an inflexible curriculum or external examination system. As Henderson et al.’s (2012) survey of pronunciation teachers found, many teachers feel that their

hands are tied by the constraints of an inflexible curriculum. As one of the surveyed teachers put it:

Spanish students need help with their pronunciation but in the end we have to be realistic.... [U]nfortunately the truth is that students must pass a written exam at the end of the year—there is no oral test. So I'm sorry to say oral skills are not the priority. (Henderson et al., 2012, p. 10)

Indeed, many English language examinations place little emphasis on pronunciation (see Chap. 6), which can have a negative motivational effect on pronunciation learning and teaching. Nevertheless, many learners and teachers recognize the fundamental importance of pronunciation in overall linguistic competence.

In the first part of this chapter, we consider pronunciation issues of teacher identity and performance, as well as teacher knowledge, beliefs, and skills, involving those charged with teaching language classes and other classes who are L2 speakers, both language teachers and international teaching assistants or assistant lecturers. In the second part of the chapter, we turn to matters of curriculum and teaching, looking first at how pronunciation can be integrated into the overall language teaching curriculum and then examining specific pronunciation teaching approaches, methods, and learning strategies in relation to research findings. The chapter closes with a discussion of implications of current views of language and language learning for pronunciation teaching practice.

Pronunciation Issues in Teacher Identity and Performance

Language Teachers Who Are L2 Speakers

Becoming a language teacher means participating in a community of practice whose members are other language teachers, both in the school and locale where the teacher's classes are located and in the larger field (Pennington, 2015; Pennington & Hoekje, 2014). Usually, a language teacher who is an L2 speaker must reference identity to more than one

language and manage identity in contexts where either or both the L1 and the L2 are salient. Developing identity as an L2 speaker and language teacher takes time. As noted in Pennington (2018):

People shape identity in an attempt to be the kind of person they want to be and imagine themselves being. Identity therefore has an aspirational aspect, an image of the self that a person wishes to project to others through word and deed, and a performed identity, which may diverge from that person's aspirational identity. The failure to perform a desired identity may be because of lack of competence or experience—the person is not skilled enough (yet) to be able to talk and act in the desired way—or lack of attention—the person is tired or distressed and so is not monitoring words and actions. There may be times when a person tries out or experiments with an alternate identity in a certain situation or with a certain audience. Over time, an alternate identity may become a more dominant part of a person's character and take on aspirational or core status. (p. 92)

Once a person has achieved the “threshold” level of language proficiency needed for effective teaching (Richards, 2012, p. 47), “[f]irst-language (‘native speaker’) competence in a language does not automatically confer [an advantage]” (Pennington & Richards, 2016, p. 11) in the way of “specific communicative skills for teaching *through* [a] language” (ibid., emphasis in original). As illustrated in Pennington and Richards (2016, p. 12), some language teachers who are L2 speakers may start off being defensive about their lack of full proficiency in the L2 but may then be able to background their status as non-native speakers as they gain teaching experience and master teaching approaches and methods. Canagarajah (1999) argued forcefully in raising awareness of the “native speaker fallacy” some years ago that L2 speakers can have an advantage in language teaching in their ability to use more than one language. They can also have an advantage in terms of relating to their students' experience, foregrounding their status as language learners and L2 speakers. As Pennington and Richards (2016) reflect: “Although ‘native speaker’ status may seem to confer an advantage in teaching a language, it has the disadvantage that a key aspect of the teacher's identity is not shared with the students” (p. 13). A suitable model for the pronunciation

of L2-speaking language teachers is not necessarily that of a native speaker but rather, that of a transcultural person or global citizen with skills in more than one language who may also share the native language of some or all of the students. The education of language teachers who are L2 speakers should incorporate this type of orientation and specifically consider ways in which those teachers can use their multiple language and cultural competences in their teaching.

As we have discussed at many points throughout this book, pronunciation, perhaps more than any other area of language, is inextricably linked to issues of identity and language attitudes: how we sound reflects who we are. This is true for teachers as well as learners; and any divergence between a teacher's accent, whether an L1 or L2 accent, and a targeted standard accent can impact the teacher's confidence or willingness to teach pronunciation. Relatively little research has focused on L1 teachers' attitudes towards teaching pronunciation with a nonstandard accent, although many teachers have a regional or nonstandard accent. The ambivalence of L2 teachers to teaching pronunciation based on their own "nonstandard" L1-influenced variety as the model for students is well documented, particularly in EIL contexts (Beinhoff, 2013; Jenkins, 2005; Henderson et al., 2012).

Jenkins (2005), for example, explored the attitudes of a group of eight teachers in terms of "their desire to identify themselves through their accents as members of an international English-speaking community" (p. 535). Jenkins describes how the teachers in her study seemed to "accept ELF in theory but not in practice" (p. 540), with views such as: "I should support EIL view as a teacher, but as a person maybe I'm aiming at native-like" (p. 540). She concludes that such perspectives reveal some uncertainties and contradictions in the teachers' beliefs between wanting to maintain their own L2 identity but at the same time feeling that teaching towards a standard native speaker model provides the best opportunity for their students. The tension between accent and identity is perhaps strongest for teachers originating from outside inner-circle countries, such as those studied by Sifakis and Sougari (2005), who were willing to consider ELF goals in theory but in practice they adhered to "inner-circle" (Kachru, 1985) pronunciation norms. Such studies illustrate how closely pronunciation is entwined with identity and social

belonging and suggest that teachers' attitudes are unlikely to change without more explicit education for pronunciation teaching and a clearer understanding of the specific social context within which pronunciation is taught. This includes increased critical awareness in different L2 teacher communities of identity, learning needs, and teaching approaches that suit their own and their students' circumstances.

International Teaching Assistants

Since the 1950s, the increased mobility and flows of people around the globe and the expanding higher education sector have meant a greatly increased number of graduate students coming from outside the country where they pursue their studies. As a way to defray expenses, these graduate students are often employed as teaching assistants or equivalent supervised assistant lecturers or course tutors. By the 1980s, university administrators and the general public recognized that these graduate teaching assistants were having significant trouble teaching students who were unfamiliar with their cultures and languages and who had very different experience and expectations of teaching (Smith, Byrd, Nelson, Barrett, & Constantinides, 1992). Many of them did not have the "threshold" level of language proficiency needed for intelligibility nor the broader communicative competence needed for effective teaching. Their lack of full proficiency and communicative competence often involved pronunciation along with other language skills.

Some universities in the United States began developing orientation programs and many different types of teaching support that were aimed specifically for **international teaching** assistants (ITAs) and that were different and more extensive than the support offered or required for other graduate teaching assistants (e.g., Altinsel & Rittenberg, 1996; Bailey, Pialorsi, & Zukowski-Faust, 1984; Nyquist & Wulff, 1996; Smith, 1993; Stevenson & Jenkins, 1994). By the late 1980s, ITA manuals and coursebooks began to appear (e.g., Byrd, Constantinides, & Pennington, 1989; Smith, Walters, & Burkhalter, 1992). By the early 1990s, 18 U.S. states required assessment of the teaching skills of ITAs using methods such as oral interviews, teaching simulations, communicative performance tests,

and some commercially produced tests (Smith, Byrd, et al., 1992). The concern about ITAs at that time and continuing to today in the United States and other English-dominant countries has fueled a considerable amount of research related to their speaking skills and pronunciation (e.g., Bailey, 1983; Bresnahan, Ohashi, Nebashi, Liu, & Shearman, 2002; Gorsuch, 2011, 2015, 2016; Orth, 1982; Pickering, 2001; Rubin, 1992; Rubin & Smith, 1990) and ongoing publication of teaching and orientation materials (e.g., Chambers, 2015; Gorsuch, Meyers, Pickering, & Griffee, 2013; Ross & Dunphy, 2007).

Research by Rubin and Smith (1990) found that undergraduate students' perception of the accentedness of ITA speech was an important variable in their assessment of their teaching but was not necessarily based on actual (or experimentally manipulated) language patterns, as the students did not always differentiate between more or less accented speech. They also found in testing students' listening comprehension that the best predictor was how many courses students had taken with non-native instructors. One of Rubin and Smith's recommendations was the need to familiarize students with non-native speech and different accents. Further research by Rubin (1992) showed that undergraduate students perceived lecturers' speech to be accented and that their listening comprehension was affected when they believed a native speaker to be ethnically Asian (based on the matched guise picture technique). A study by Orth (1982) showed that undergraduate students' assessments of ITAs' speaking skills often did not match those of expert raters and were biased by their expectations of the grades they would receive from the ITAs, that is, by how difficult they perceived the teacher and the course to be. A study by Bresnahan et al. (2002) suggested that undergraduate students judged ITAs more positively when they found their speech accented but intelligible, but also judged them more negatively in terms of accented speech than they did student peers. Pickering (2001) pinpointed intonation as a key to effective native speaker TA performance and a problem for a matched group of ITAs, who were not able to successfully manage intonation for indicating information structure and establishing rapport with students.

Intonation and other prosodic aspects of speech are important aspects of teaching which ITAs need to master, as addressed in the early course

materials (e.g., Byrd et al., 1989; Smith, Walters, & Burkhalter, 1992) as well as more recent materials and courses aimed at improving ITA performance (e.g., Chambers, 2015; Gorsuch et al., 2013; Ross & Dunphy, 2007). Gorsuch's research agenda on ITAs has resulted in a number of published studies focused on ITA training. Gorsuch (2011) developed a technique of having ITAs repeatedly and silently read popular science texts with reference to audio recordings of those texts, considered to be a form of comprehensible input. The technique improved the ITAs' reading comprehension and reading fluency, in addition to their production of pause-groups, as an indicator of improvement in their meaning-related information-grouping and pausing behaviors. Gorsuch et al. (2013) developed an in-depth course for ITAs that includes audio and video recordings of classes as well as a range of skill-building and interactive activities working on information structure, including thought groups and "speech paragraphs"; on discourse intonation and the use of rising and falling tones for various communicational functions, including for establishing rapport and interacting with students; and on spoken grammar, intelligible pronunciation, and fluency. The Gorsuch et al. (2013) book includes information for ITA course instructors on assessment of ITA performance, which is also commonly assessed using standardized tests of spoken language (as described in Chap. 6), especially Pearson Longman's *Versant* test (Chaps. 5 and 6).

A working paper published online (Gorsuch, 2017) offers a detailed ITA evaluation procedure focused on discourse intonation and speaking strategies and suggests its value in improving various aspects of math and science ITAs' discourse intonation and speaking strategies, including information management and grouping, question and answer elicitation and response patterns, and word-level pronunciation focused on repetition and prominence as communication strategies. Gorsuch's working paper, together with her co-authored book and published articles, can be considered to represent the state-of-the-art in ITA research and training. The work of Gorsuch and colleagues thus offers useful directions and guidance for ongoing research on areas of pronunciation that are most essential for ITA teaching effectiveness and how best to remediate pronunciation problems that are also suggestive for improving pronunciation and related communication skills of any L2 speaker, including language teachers.

Teacher Beliefs, Knowledge, and Skills Related to Pronunciation

Surveys have confirmed that much less classroom time is devoted to pronunciation than to other areas of language learning, despite the importance that many learners place on pronunciation and the fact that research shows that more time spent pays off in better results (Lee, Jang, & Plonsky, 2015). Foote et al.'s study (2013) of teachers of three grade 6 ESL classes in Québec, for example, suggested that about 10% of class time is given to pronunciation at grade 6 level, compared with 20% for grammar and 70% for vocabulary. It is very likely that these proportions of time devoted to different aspects of language represent a general understanding within the field of language teaching as to their relative importance, together with considerations of what is most teachable. There may be other reasons for the limited focus on pronunciation within language teaching as well, including, for example, lack of teacher confidence, skills, or knowledge regarding pronunciation teaching; lack of appropriate teaching materials; or lack of time or space in the curriculum. Until pronunciation is given greater priority in language teaching and testing, the situation is unlikely to change substantially, although there are signs for optimism, with growing awareness of pronunciation in some high-stakes tests (Chap. 6) and some syllabuses and textbooks giving considerable space to pronunciation. At present, however, many language teachers feel underprepared, in terms of knowledge and skills, to teach pronunciation and recognize a need for more systematic professional development and better teaching resources to support them (Henderson et al., 2012; MacDonald, 2002; Thomson, 2011).

As a result of this lack of training and resources, teachers may have to develop methods based on intuition or old-fashioned (outdated) pronunciation teaching materials, rather than evidence of what is most effective. At times this may work and teachers can find ingenious ways of explaining or demonstrating pronunciation features, such as word stress, vowel length, or the pronunciation of individual segmental phonemes. In addition, teachers who are experienced and observant are well placed to conduct their own classroom-based action research to see what works best

with their students. However, there are dangers inherent in adopting an intuition-based rather than an evidence-based approach to pronunciation teaching, particularly in terms of teaching based on inaccurate knowledge (e.g., “an English /b/ is produced while breathing in,” quoted in Thomson, 2013, p. 229), but also in terms of spending time teaching features that may not be priorities, such as the /θ, ð/ contrast for speaking in ELF contexts (Jenkins, 2000) or teaching to a specific regional or national model for pronunciation that does not match learners’ needs in terms of speakers they are likely to encounter on a regular basis (e.g., teaching to a British pronunciation model in the Americas). Derwing and Munro (2005) suggest that “teachers show a great deal of confusion about what is possible and what is desirable in pronunciation instruction, despite some quite definitive research findings on this topic” (p. 384), though we maintain that there is in fact little truly “definitive” research, given the many factors that can be considered to have an impact on instruction in any given context. The issue of intuition-based versus evidence-based teaching and the gap between research and pedagogy is discussed in more detail in Chap. 8.

Integrating Pronunciation with Other Areas of the Curriculum

Given the various constraints on instructional time and content, language teachers need to be as resourceful as possible in finding ways to integrate pronunciation into the language learning syllabus. One way of doing this is by incorporating pronunciation into other areas of language learning, and the benefits of doing so have been highlighted by a number of educational practitioners (e.g., Celcia-Murcia, Brinton, Goodwin, & Griner, 2010; Dalton & Seidlhofer, 1994; Rogerson-Revell, 2011).

Speaking and Listening

Opportunities for integrating pronunciation into general oral skills development are fairly well documented (e.g., Levis & Grant, 2003; Murphy, 1991)

and increasingly part of the education of language teachers. Work on pronunciation, whether in the form of explicit lessons or as feedback on speaking performance, can be a part of all kinds of oral work, including conversational and other interactive speaking tasks as well as structured communication tasks of oral presentation, speech-making, and debating. Work on the pronunciation of individual sounds can improve the intelligibility of individual words, as can work on stress in multisyllabic words and phrases. Work on stress and other aspects of prosody, such as intonation and coarticulatory linking or elision, can improve comprehensibility more generally as well as fluency, utterance pragmatics, and audience rapport and impact.

Pronunciation-focused instruction can aid learners' listening comprehension, in being able to reliably identify minimal pair words as well as words and phrases with reduced vowels and other forms of phonological reduction (e.g., contraction and cluster simplification). Cloze passages can be designed to focus specifically on contractions or on function words that are typically greatly reduced in rapid speech, such as articles (*a, the*), conjunctions (*and, but, or*), and short prepositions (*of, on, in, at*). A focus on prosody might include listening work to identify pitch contours, stress placement in words and sentences, and differences in meaning associated with these, including pragmatic or interpersonal meaning. A wide range of activity types are possible, such as marking of stress in multisyllabic words or contrastive stress in sentences; identification or discrimination of minimal pairs; marking of intonation as falling, rising, or neither; or marking the correct pronunciation of plural, possessive, and past tense endings based on final consonant. (For further examples of listening activities focused on pronunciation, see Rogerson & Gilbert, 1990, pp. 107–130; Celce-Murcia, Brinton, & Goodwin, Griner, 2010, pp. 370–393.)

Besides aiding listening comprehension, work on pronunciation within listening tasks can aid perception in ways that are expected to have “wash-back” for speaking, building awareness of phonological features and cues as advocated by Couper (2015). Pronunciation fluency and accuracy practice can be introduced into speaking and listening lessons, through activities such as student-to-student dictations, creating and performing

news commentaries or weather forecasts, and recording and writing down answerphone messages (Rogerson-Revell, 2011, pp. 252–253). Muller Levis and Levis (2016) describe many ways in which prosodic aspects of pronunciation can be integrated into the teaching of oral skills. In addition, speaking and listening lessons can aim to raise learners' awareness of pronunciation features related to different registers or styles of speaking (e.g., formal vs. informal) appropriate to different kinds of communicative tasks and purposes, as well as to different varieties (e.g., London vs. New York City, or GB vs. GA vs. EIL). With an eye to learners' communicative needs, teachers can survey their students about whom they expect to be interacting with most in future and include attention to the pronunciation features of those groups. With the aim of focusing learners' attention to the maximum extent on sound and moving them away from L1-influenced sound-spelling correspondences, we recommend that teachers seek out listening and speaking practice activities which can be performed without reliance on the written language, or adapt textbook activities so that learners are required to close their books for some phase of listening/speaking activity.

Grammar and Vocabulary

Teachers can also be encouraged to integrate pronunciation into other areas of language, such as grammar teaching and vocabulary development. For example, intonation patterns that are associated with different grammatical patterns and utterance functions, which vary with language variety and context, are an important aspect of communicative knowledge that might be taught in relation to teaching the grammar of clauses, sentences, and larger discourses. Intonational contrasts between new or focal and given or backgrounded information are also important aspects of grammar taught in relation to communicative function for purposes of both comprehension and production of speech. Communicative tasks can be designed to show these contrasts in phrasal, sentential, and discourse contexts (e.g., see Pennington, 1996, exs. 2 and 3A, 4B, p. 177–181). In addition, students' awareness of prosodic features such as

contractions, weak forms, and stress in complex conditional clauses can be raised, such as by having them listen to and then mark contractions, stress, and linking in examples such as (e.g., “If I’d known about the party, I’d’ve come”) (see Rogerson & Gilbert, 1990, pp. 31–37).

The teaching of grammatical forms that tend to be phonologically reduced, including auxiliaries (forms of *be* and *have*) and verbal and nominal inflections (in particular, *-ed* / *-en*, *-(e)s*, *-ing*), can usefully include instruction on pronunciation to aid perception and production. If students fail to produce these inflections, the reason may be a failure to notice them in perception of everyday speech. Working on the difference in pronunciation between the stressed and unstressed syllables in noun/verb pairs spelled the same (e.g., *object*, *project*, *overlap*) and sets of words formed from the same root that have variable stress placement and phoneme changes in vowels and sometimes also in consonants (e.g., *pronounce/pronunciation*, *explain/explanatory/explanation*; *photo/photograph/photography/photographer*; *resign/resignation*), such as by having students mark word stress and identify both full and reduced vowels, will also be an aid to perception and production of these in speech.

The pronunciation component involving vowel shift in the singular and plural of irregular nouns (*tooth/teeth*, *foot/feet*, *man/men*, *mouse/mice*), and in the present and past tense and participial forms of irregular verbs (*see/saw/seen*, *come/came/come*, *know/knew/known*, *feed/fed/fed*, etc.), might be pointed out in teaching those irregular but common nouns and verbs to aid memorization as well as perception and production in listening and speaking. Other pronunciation patterns in grammatical morphology that can be pointed out as an aid to learning and perception/production are those involving voiceless and voiced final fricative in singular/plural pairs with /f, v/ (e.g., *wife/wives*, *life/lives*, *self/selves*) or singular/plural and noun/verb pairs with /θ, ð/ (*bath/bathes*, *wreath/wreathes*, *breath/breathes*). Celce-Murcia et al. (2010, pp. 397–418) discuss these and other aspects of grammatical morphology that have pronunciation components which can be part of instruction, and they offer several examples of classroom exercises focused on these features.

Literacy and the Written L2

Language learners often do not begin learning a second language until both pronunciation patterns and sound-spelling correspondences are well-established in their L1. Thus L1 transfer (Chap. 2) in the spoken language often involves transfer of L1 phonological categories and articulatory patterns as well as L1 orthographic patterns tied to pronunciation. It can therefore be of value to explicitly teach the orthographic patterns of English and how these differ from those of other languages which learners know. It may also be useful to teach new words first by sound and only relate these to their written form once their pronunciation has been established, such as through listening and repetition.

In addition, many languages contain a very large vocabulary of English-based loanwords with their own conventional patterns of pronunciation and sound-spelling correspondences, including adjustments of syllable types to those of the L1. Low-level learners may base their pronunciation of English words on the corresponding L1 loanwords, and only gradually move away from an L1 loanword to the L2 lexical item and its pronunciation according to L2 conventions. Explicit contrast of L1 loanwords with L2 pronunciation of those same lexical items may be of value (Rogerson-Revell, 2011, p. 214), especially for low-level learners. If the teacher is not familiar with those loanwords, it can be a useful lesson in contrasting the L1 and the L2 for the students to supply them and to pronounce them in the L1 form, and then for the teacher to model the pronunciation of the corresponding L2 word.

Many L2 speakers are already literate and often experienced readers when they begin acquiring a second language, so they may tend to pronounce L2 words based on L1 pronunciation as tied to L1 orthographic patterns, as an aspect of L1 transfer. This is especially likely when L2 instruction focuses more on reading and the written language than on oral performance and the spoken language. Learning a language from the written language or with a heavy emphasis on reading often results in “spelling-pronunciation” errors, in which words are pronounced according to speakers’ assumptions about the sound-value of graphemes, such as

pronouncing the “silent *t*” in *often*, or pronouncing L2 words according to L1 sound-spelling correspondences. According to Koda (2007), “Reading transfer studies collectively demonstrate that L1 orthographic experience has clearly detectable and long-lasting impacts on the formation of decoding skills in the L2, and thus constitutes a major source of performance variation in phonological processing among L2 readers” (p. 233). Sound-focused listening (see below), such as to different speakers’ voices and different varieties, and listening without the benefit of a written script, can help to minimize L1 transfer. In addition, learning the phonetic alphabet and using it as a reference for listening to and speaking the L2 can move learners away from a reliance on L1 pronunciation and sound-spelling correspondences. It may also be of value to have students mark or otherwise identify silent letters in new vocabulary (e.g., *knee*, *thumb*).

The Value of Integrating Pronunciation into Other Areas of Language Learning

Integrating pronunciation teaching into other areas of language learning is a way of reminding learners of its broader significance. It is worth spending some time conveying this awareness and concern for pronunciation to learners as early as possible in the learning process. Research has suggested that the greatest gains and improvement in pronunciation can be made in the first twelve months of instruction or exposure to the target language (Munro & Derwing, 2008), and the early stage of L2 learning is especially valuable for pronunciation if learners are immersed in real-language input. If students can be convinced of the broader impact of pronunciation on communication, not just in terms of accuracy and intelligibility, but also in relation to fluency and impact, they are more likely to be motivated to improve. Also, if the teacher can be seen to be taking learners’ needs into account in setting goals and deciding content, it is likely to further increase motivation. We consider in more detail in Chap. 6 how a teacher can informally evaluate learners’ pronunciation and diagnose problem areas.

Teaching Approaches for Pronunciation

Order and Focus of Instruction: Bottom-Up vs. Top-Down, Segmental vs. Prosodic

Traditionally, pronunciation materials or syllabi start with the smallest elements, that is, individual phonemes, and work up towards larger elements such as word stress, features of connected speech, and intonation. However, such a linear approach can make it hard for learners to get the bigger picture or to understand how the various elements fit together or interrelate. Alternatively, the teacher can start by introducing prosodic aspects of pronunciation as key to production and perception of speech. The teacher might, for example, start by explaining that the syllable is central to pronunciation and that the combination of syllables, each containing a vowel, creates the typical rhythmic patterns of English with its alternating stress–unstress pattern. A connection could also be shown between the rhythmic structure of English and vowel quality, and the fact that stressed vowels are clearer and longer than unstressed vowels, which are typically reduced to schwa. This, in turn, could make learners aware of the need to hear and produce stress patterns accurately. A general introduction to English rhythmic patterns can also demonstrate the influence of stress on phrasing and coarticulation.

A related issue regarding the organization of pronunciation teaching content is what to emphasize, whether to adopt a traditional, bottom-up or micro-level phoneme-based segmental orientation, possibly as the exclusive approach or possibly as an initial stage leading to work on prosody, or a top-down, macro-level orientation emphasizing suprasegmental or prosodic aspects such as rhythm, prominence, and intonation, as advocated in Pennington (1989), either as the initial or only focus of pronunciation instruction. Some researchers (e.g., Levis, 2005; Saito, 2014) suggest that segmental features may be easier to teach and learn, while others claim that focusing on suprasegmentals is more effective (Hahn, 2004) as well as more efficient in the sense that work on prosody also impacts segmentals (Pennington, 1989), as shown, for example, in the work of Hardison (2004). As in all teaching decisions,

teachers must always consider whether what is easiest to teach and learn should have priority over other instructional content that could have more overall benefit or “payoff” in the end. Although the jury is still out as to the relative benefits of a teaching focus on segmental versus suprasegmental aspects of pronunciation, including connected speech phenomena having to do with phonological fluency and prosody, Lee et al.’s (2015) meta-analytical review of pronunciation teaching effectiveness research showed a greater effect when both segmental and suprasegmental features are included in instruction.

We have seen earlier that there has been a historical shift in approaches to pronunciation teaching, with an increasing focus in teaching materials since the 1980s on teaching prosody but now a resurgence of attention to segmentals in the “focus on form” orientation of task-based and other communicative teaching approaches. Research gives some support to a focus on prosody, particularly for comprehensibility with L1 listeners. From their research, Derwing and Munro (1997) concluded that “improvements in NNS comprehensibility, at least for intermediate and high-proficiency learners, is more likely to occur with improvement in grammatical and prosodic proficiency than with a sole focus on correction of phonemic errors” (p. 15). Studies have shown that L1 listeners pay more attention to prosody than to phonemic accuracy (Leather, 1999) and that prosody has a greater influence on pronunciation ratings for standardized spoken language tests (Anderson-Hsieh, Johnson, & Koehler, 1992). Kang, Rubin, and Pickering’s (2010) study using samples from the iBT TOEFL Practice Online test (see Chap. 6) found that, with L1 raters, “suprasegmental features alone can collectively account for about 50% of the variance in proficiency and comprehensibility ratings” (p. 564), and Isaacs and Trofimovich (2012) discovered in their study of comprehensibility that word-stress errors were a key feature differentiating learners at all levels of proficiency. Other research shows that correct sentence intonation is one of the best predictors of speaking proficiency (de Jong, Steinel, Florijn, Schoonen, & Hulstijn, 2012) and that prosody is an important feature differentiating L2 proficiency level of speakers (Long, Gor, & Jackson, 2012).

Most of these studies were conducted with L1 listeners or raters while in her study of communication breakdowns in interactions between L2

English speakers, Jenkins (2000) found that the majority of communication breakdowns were due to segmental errors, or segmental combined with nuclear stress (i.e., sentence- or utterance-level stress) errors, leading her to conclude that “for EIL, and especially for NBESs [nonbilingual English speakers], the greatest phonological obstacles for mutual intelligibility appear to be deviant core sounds in combination with misplaced and/or mispronounced nuclear stress” (p. 155). Low-level learners were found to have markedly more difficulties of this kind than more advanced learners. These findings can be compared to those of a major study by Goh (2000) showing that phoneme and word recognition are a major source of difficulty for low-level L2 listeners.

It seems that learners with a low level of proficiency tend to process speech largely based on bottom-up rather than top-down information processing strategies (Field, 2003; Wilson, 2003); in other words, they focus attention more on individual sounds and words than the higher level and more global units of prosody, other aspects of linguistic context, and situational context. Such a focus may reflect the extra attentional demands and cognitive load of processing L2 speech, which is less automatized than the processing of speech in the native language (Chap. 2). Jenkins (2000) makes a related point regarding learners in multilingual settings, suggesting that they rely more heavily on the acoustic signal, which “in turn diverts cognitive resources away from features of the context, which are thence not available to compensate for any limitations in speech perception or production” (p. 83). This creates a vicious circle in the sense that learners focus so much on bottom-up, moment-to-moment processing of the acoustic signal that they miss cues to meaning in longer stretches of speech and in the wider context due to cognitive overload.

It would seem that further research is needed into the processing of speech by L2 speakers and listeners in regard to the relative value of more micro- or macro-focused pronunciation teaching. In the meantime, a conservative position is that some aspects of both segmental and suprasegmental pronunciation, and of bottom-up and top-down approaches, are important, depending on the teaching context and learners’ needs. Perhaps the key point is to show that the various features of pronunciation are not discrete elements but interact to facilitate fluency and intelligibility. As Zielinski (2015) says, “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” (p. 407). One way that teachers can do this, for instance, is by including work on rhythm and connected speech which shows how individual sounds are modified in the stream of speech and also shows the interrelationships of rhythmic patterning, word stress, intonation, and informational prominence.

A different sense of top-down or macro-level teaching is to focus first on broad functions of language having to do with pragmatics and communicative framing or purpose through discourse structure and global features of prosody (e.g., **discourse intonation**; Brazil, Coulthard, & Johns, 1980) before focusing on the bottom-level details of linguistic form, including pronunciation of specific lexis and grammatical morphology as these are sequenced in real time. It can be noted that this order of focus matches that which Levelt (1989, 1999) suggests is the order in which communication proceeds when a person prepares to speak, starting at the most global or macro level of discourse planning and then developing plans at each succeeding more micro level that results in the detailed mechanics of sequential articulation (Chap. 2). Such an approach could be applied, for instance, in preparing to perform a textbook dialogue or in planning, rehearsing, and then performing a roleplay or a speech, with the student creating a specially marked up script showing phrasing, intonation curves, stress, and contextual assimilations and reductions—which is something that public speakers and actors sometimes do to help them develop and practice their desired delivery.

Another sense of broad-based or macro-focused teaching would provide large doses of listening input, that is, **extensive listening**, at an early stage of language learning, as a way to build the learner’s internal database of L2 sound. Such “input-heavy” approaches might for instance flood learners with different kinds and varieties of natural speech input in the early stages of language learning, possibly through both extensive and intensive (focused) listening and with some sound-based imitation of multiple perceptual models. These approaches aim to imprint the sound patterns of the language from the outset through implicit and explicit learning processes (as detailed in Chap. 2) before those patterns are tied to specific

grammatical structures or written language forms that will initiate L1 transfer. This is not the same type of localized “input-flooding,” focused on an individual form or a contrasting set of forms, advocated for language teaching by Doughty and Williams (1998). It is rather a global type of input-flooding, providing an intentionally varied and diversified flood of L1 speech, including multiple samples of discourse, potentially in forms ranging from monologue to dialogue, storytelling and reading aloud, to famous speeches and conversations, and potentially incorporating different people’s pronunciations of specific words, phrases, and sentences as well. Samples of these kinds are readily available and can easily be found online.

The point is to begin to build the learner’s awareness and cognitive representations of L2 speech, based on a massive amount of input, before undertaking organized instruction, and especially before using the support of the written language.¹ Other than this input-heavy listening approach, it can also be of value to teach strictly through the spoken language at the initial stage of language learning, as a way to ensure that learners will have to listen carefully in order to perform tasks and thus imprint the pronunciation patterns of the language outside of any connection to their knowledge of L1 pronunciation patterns, including those related to lexical and grammatical patterns which become salient once written words are viewed. Some basic vocabulary and question-and-answer patterns, for example, can be introduced and practiced in the first week of a language class without any use of written aids, and then these can be built on with further vocabulary and patterns in the weeks that follow. It can be of considerable value for students’ future listening and speaking, as a way to ensure that they will be focused on the details of sound, to spend at least the first several hours of instruction in this way. An exploratory study of the teaching of Turkish in an optional university class (Pennington & Körmücü, 2018) in which the first 6 hours were taught in this way demonstrated the value of this approach for sharpening students’ attention and listening skills, resulting in both rapid communication-focused learning and pronunciation that was highly imitative of the L1 Turkish teacher’s speech. At the same time, the exploratory study made clear the discomfort which the teacher and some of the

learners (especially a Chinese student who was an L2 English learner) experienced in not having any written support for learning.

The Limits of Meaning-Focused Instruction for Fine-Tuning Pronunciation Output

The latter two approaches—input-heavy listening and the exclusively listening–speaking orientation of Pennington and Körmücü’s (2018) exploratory curriculum for teaching Turkish—both aim to build students’ awareness and cognitive representations focused on the sound of an L2. These approaches thus contribute to a long line of teaching approaches that have aimed to help students learn language based on input that will be naturally absorbed and processed by the learner’s cognitive mechanisms, including Communicative Language Teaching (Brumfit & Johnson, 1979; Widdowson, 1978), the Natural Approach (Krashen & Terrell, 1983; Terrell, 1977, 1982), and Task-based Language Teaching (R. Ellis, 2003; Long, 2015). In Krashen’s (1982, 1985) view, as adopted and adapted in SLA theory more generally (e.g., Long, 1981, 1983), language acquisition takes place in communicative interactions when a learner is focused on meaning and is exposed to language that is just beyond the learner’s current level of knowledge or competence but that is made understandable through contextual information. The key to acquisition of spoken language is a learner’s exposure to such **comprehensible input** within a relaxed and receptive frame of mind, free of fear of speaking or anxiety towards the L2 or its speakers—a state described by Krashen (1982, 1985) as one in which the learner’s “**affective filter**” is low. According to Krashen’s (1982, 1985) **Monitor Model**, conscious learning can be applied in communication situations to monitor, adjust, or correct performance by a momentary focus on correctness or output form. For children learning their mother tongue, conscious learning should not be necessary for acquiring spoken language competence, as they would learn to speak their L1 perfectly through unconscious acquisition (i.e., implicit learning). For adults learning an L2, however, assuming that acquisition mechanisms would not apply as effectively as in L1

acquisition, conscious learning would be available to improve spoken language performance (see also Chap. 2).

In the 1970s and 1980s, the focus of language teaching was on providing input with contextual supports to make it comprehensible, limiting attention to explicit teaching or to learners' errors, in an effort to allow natural acquisition processes to operate. However, it was soon realized that a focus on communicative interaction often resulted in learners who, while they acquired language and developed communicative skills to a certain level, nonetheless had various deficiencies in knowledge of the L2 and communicative competence, including fossilized error types (Higgs & Clifford, 1982) and, in many cases, relatively pidginized speech (Klein & Dimroth, 2009; Schmidt, 1983; Schumann, 1978). As concluded by Long (2015): "Adult [L2 acquisition] can be done incidentally and implicitly up to a point, in other words, from exposure to, and communicative use of, the L2, but the results always fall short, usually far short, of native-like abilities..." (p. 38). The input that learners were receiving in communicative and meaning-oriented approaches to language learning and teaching were not resulting in sufficiently detailed, correct, or "well-tuned" output to achieve an advanced level of communicative competence and in some cases, fell considerably short of even a basic level of intelligibility or comprehensibility. The focus on meaning and communication was apparently not tuning perception sufficiently to **input form** and so not establishing a basis for tuning production sufficiently to **output form**.

Communicative competence incorporates not only general communicative abilities making it possible for learners to get their meaning across, but also the ability to fine-tune their output to achieve high accuracy, clarity, and specificity of meaning, pragmatic force, and social impact in different circumstances. Although it is widely accepted in SLA that instruction needs to have a basis in communicative and meaning-focused learning, it has long been clear that such an orientation is not sufficient. In a state-of-the-art review of CLT published some years ago, it was already possible to assert that "for the development of communicative ability, research findings overwhelmingly support the integration of form-focused exercises with meaning-focused experience" (Savignon,

1991, p. 269). Since the 1990s, attempts to integrate a focus on form within communicative and meaning-oriented instruction have had a major impact in the teaching of pronunciation.

Form-Focused Pronunciation Instruction

The re-emergence of interest in micro-level, segmentally focused pronunciation teaching parallels to a considerable extent recent interest in cognitive approaches to pronunciation learning, such as form-focused instruction (FFI), to draw learners' attention explicitly to target language features that they would otherwise not notice (Spada, 1997). Tarone, Bigelow, and Hansen (2009, pp. 29–30) review the *cognitive comparison process* described by SLA researchers and theorists that is required for corrective feedback to result in uptake of new features into an L2 learner's developing linguistic system and performance. The process has the three steps of (1) noticing the correct form as given in the feedback, (2) comparing that form in working memory to the learner's own incorrect form, and then (3) incorporating the correct form into the learner's performance and linguistic system in long-term memory. It is thus a type of explicit learning focused on form. Long (1991) introduced the idea that a focus on form was a design feature of language teaching. Since then focus on form has become an important consideration in language teaching design and research (Doughty & Williams, 1998; Wong, 2005). "Form focused instruction (FFI) attempts to encourage noticing, drawing learners' attention to linguistic forms that might otherwise be ignored" (Cintrón-Valentín & Ellis, 2016). These are forms which learners ignore either because of their low functionality and cue value (i.e., their low salience, low contingency, and redundancy; Wulff & Ellis, 2018) or because of transfer of entrenched L1 patterns of perception and production (resulting in L1-tuned learned attention, blocking, and automatized processing; Wulff & Ellis, 2018), or both.

While teaching approaches with an emphasis on error-correction or accuracy had been dismissed in the 1970s and 1980s in the light of communicative approaches to language teaching, the value of FFI has increasingly been recognized, as supported by cognitive theories of SLA

and phonology (Couper, 2015; Mompean, 2006, 2014). The idea that learners need to pay attention to and notice target language input and understand the significance of noticed input is central to many aspects of contemporary SLA theory (Cintrón-Valentín & Ellis, 2016; Long, 2015; Schmidt, 2001; see Chap. 2 for discussion). The importance of explicit instruction and the centrality of concepts such as noticing and attention, which are well established in other areas of language learning such as grammar, are gaining ground in pronunciation teaching and learning.

R. Ellis (2009) divides FFI into implicit and explicit types:

Implicit FFI

- Attracts attention to target form
- Is delivered spontaneously (e.g. in an otherwise communication-oriented activity)
- Is unobtrusive (minimal interruption of communication of meaning)
- Presents target forms in context
- Makes no use of metalanguage
- Encourages free use of the target form

Explicit FFI

- Directs attention to target form
- Is predetermined and planned (e.g. as the main focus and goal of a teaching activity)
- Is obtrusive (interruption of communicative meaning)
- Presents target forms in isolation
- Uses metalinguistic terminology (e.g. rule explanation)
- Involves controlled practice of target form (p. 18)

Explicit FFI is a kind of consciousness-raising through explicit teaching and learning aimed at improving performance. Implicit FFI in contrast aims to improve performance through implicit teaching and learning, in what may be regarded as “subconsciousness-raising.”

Implicit FFI may be focused on quantity of input, through *input-flooding* (Doughty & Williams, 1998) involving a range of ways to produce a flood of input, from providing a discourse in which multiple repetitions of a language form (a sound, word, or phrase) naturally occur in one communicative context to providing multiple communicative contexts in which a specific language form occurs. Or, implicit FFI may be focused on input quality or variety, such as through use of input from different dialects or varieties of the L2, different ages and genders of speakers, or more careful versus informal pronunciation. In the context of phonology, it has been specifically recommended for input to be geared to variety, as in Esling's (1987) notion of "collection not correction" (p. 469) and in the concept of **high variability perceptual/phonetic training** (HVPT) as first developed by Logan, Lively, and Pisoni (1991).

Explicit FFI may likewise be focused on quantity of input, such as an explicit focus for perception and/or production on multiple repetition, or it may be focused on quality of input, such as different forms of input enhancement (Wong, 2005),² or selection of specific exemplars of a form that are judged to be typical or otherwise of good quality. Input enhancement may incorporate various kinds of visual enhancements, such as to indicate stress patterns in multisyllabic words and in sentences, and aural enhancements by computer (see Chap. 5). Explicit FFI might include:

- **pre-input** aimed to focus attention or guide performance in certain ways, such as through giving a rule or procedure to follow, modeling, or preview activities that provide opportunities to practice within a limited context;
- **task-oriented input** (termed "task-essentialness" by Doughty & Williams, 1998) and form-focused aspects of performance of instructional tasks, such as Muller Levis and Levis' (2016) "bridging exercises" bridging between form and meaning; and
- **post-input**, that is, output-focused input, such as corrective feedback, post-task teaching of various kinds, and post-reflection.

Pronunciation awareness-raising tasks can be recommended for pre-input and task-oriented input aimed at preparing learners for successful

task performance as well post-input (feedback) on completed tasks—most effective if based on video or audio recordings—with the aim of sharpening perception and future performance. Similar awareness-raising tasks can be developed for the prosodic aspects of speech, such as stress patterns in lexical items essential to accomplishing a specific communicative purpose. In addition, explicit FFI may involve opportunities to practice the mechanics of producing sounds and prosodic features such as stress in isolation or in context for better discrimination and higher intelligibility.

Long (2015) restricts focus on form to post-input, that is, input such as recasts that target the form of output. For him focus on form is “by definition *reactive*, not *proactive*” (p. 316 n. 7, emphasis in original). Long (1981, 1983) proposed that brief periods of FFI could be especially effective during negotiation for meaning, drawing the learner’s attention to form (e.g., through corrective feedback or a clarification request) in order to facilitate a meaning-oriented goal. In this way: “Intentional learning is brought to the aid of incidental learning, thereby improving the likelihood that a new form-meaning association will be perceived or perceived more quickly” (Long, 2015, p. 317). In his Interaction Hypothesis, Long (1996) notes that “*negotiation for meaning*, and especially negotiation work that triggers *interactional* adjustments by the NS or more competent interlocutor, facilitates acquisition because it connects input, internal learner capacities, particularly selective attention, and output in productive ways” (pp. 451–452, emphasis in original). Focus on form within negotiation for meaning, such as in Communicative or Task-based Language Teaching, “utilizes the learner’s capacity for incidental learning, which will be necessary, due to both the size and complexity of learning task, while improving on pure incidental learning through systematic recruitment of intentional learning, but doing so reactively, in harmony with the learner’s developmental readiness” (Long, 2015, p. 321). Other kinds of FFI that are not in response to specific acts of communication by the learner are what he terms “focus on forms” (in the plural).

A number of studies have shown positive effects of focus on form in Long’s sense as well as focus on forms. A meta-analytical study by Norris and Ortega (2000) confirmed positive effects for both, with stronger

effects for focus on form than focus on forms and strong effects for explicit versus implicit instruction. In another meta-analytic study, Spada and Tomita (2010) report larger effect sizes for explicit than implicit instruction and suggest that the explicit instruction learners received affected not only their controlled knowledge but also their usage of the forms that were the focus of instruction. A third meta-analytic study by Goo, Granena, Yilmaz, and Novella (2015) also found a greater effect for explicit than implicit instruction, though that for implicit instruction was significantly larger than what Norris and Ortega (2000) had found. As Long (2015) observes, many of the studies of explicit instruction had a primary focus on “code features” and so, although effective in teaching those code features, do so “at a major cost” (p. 320), losing the value of implicit or incidental learning for the monumental learning task that language learning represents. Focus on form can be seen as “an attractive alternative” (ibid.) that recruits a learner’s capabilities for intentional learning from explicit instruction in the service of the incidental language learning which takes place when the primary focus is on something else, such as a body of knowledge that a learner wishes to learn or a communicative task that the learner wants to perform.

The mechanism of an interface making it possible for explicit input to become implicit knowledge is not known. While some researchers argue that interface is not possible, others argue that it is. N. C. Ellis (2005) argues that “implicit and explicit knowledge are dissociable but cooperative” (p. 305, Abstract), so that “explicit knowledge of form-meaning associations [can impact] upon implicit language learning” (ibid.). As he describes it, “The interface is dynamic: It happens transiently during conscious processing, but the influence upon implicit cognition endures thereafter” (N. C. Ellis, 2005, p. 305, Abstract). A “weak interface” position is accepted by many researchers in SLA through which “...explicit knowledge [is seen] as contributing indirectly to the acquisition of implicit knowledge by promoting some of the processes believed to be responsible” (R. Ellis, 2009, pp. 21–22).

Both R. Ellis (1994) and N. C. Ellis (2005; Ellis & Wulff, 2015) have proposed that consciously applied **declarative knowledge** can help learners notice features in the input which they otherwise might miss. Once a feature of input has been noticed, it becomes intake to implicit processes

for pattern recognition that tune future perception and production. This is one of the ways N. C. Ellis (2005) describes in which explicit learning can dynamically “cooperate” with implicit learning processes. In addition, as R. Ellis (2009) points out, the dividing line between explicit and implicit learning is in many cases not clear, and, no matter which type of FFI is intended, the learner may learn in an explicit or implicit way, since both implicit and explicit learning can occur as a result of exposure to input in the context of instruction rather than to the instructional presentation:

[A] learner can always elect to respond to what the teacher says as ‘input’ rather than as ‘information’. In such a case, explicit instruction can result in implicit learning as a result of the incidental noticing of instances of language. Equally, in the case of direct intervention involving implicit instruction, learners may work out what the target of the instruction is and seek to make their understanding of it explicit. Thus, it does not follow that implicit instruction always results in implicit learning or that explicit instruction necessarily leads to explicit learning. It should also be noted that the aim of explicit instruction is not just to develop explicit knowledge but also, ultimately, implicit knowledge as well. (p. 18)

As a final point, it can be suggested that the sharp division between implicit learning and explicit learning—both of which might be triggered by instruction—is not of great significance, once it is agreed that all learning involves general cognitive learning processes requiring noticing and attention, as Schmidt (1990, 1992, 2001) argued, and analysis of input, as N. C. Ellis (1996, 2002, 2005) has described.

In relation to teaching individual phonemes, perception-based phonetic training has long been promoted as essential to establish the foundations of pronunciation. Research has shown that perceptual training tasks, such as phoneme discrimination and identification tasks, can be particularly effective and can also facilitate the development of productive skills (Bradlow, Akahane-Yamada, Pisoni, & Tóhkura, 1999; Thomson, 2011). Several recent studies have shown that HVPT, as a particular approach to perception-based phonetic training, is highly effective in enabling learners to improve their ability to discriminate

linguistically contrasting sounds (see Thomson, 2011; Wang & Munro, 2004; Wong, 2015). Unlike traditional approaches to pronunciation teaching, which typically focus on a single, standard model, HVPT emphasizes the use of multiple speakers and diverse phonetic contexts to increase learners' awareness and tolerance of variation. Some studies have further shown that the training effect can be sustained over time (Bradlow et al., 1999; Lambacher, Martens, Kakehi, Marasinghe, & Molholt, 2005) and is also effective at different levels of proficiency (Wong, 2015).

The point that many of these studies highlight is the need for language learners to learn how to listen accurately, that is, to notice specific phonological features in the stream of speech and also to understand how and why they are important. This means they need to attend to pronunciation and become consciously aware of their contribution to speech in their own and other speakers' performance. This way of approaching pronunciation is quite different from the drilling approach of form-focused instruction in the structural linguistics and audiolingual period, as it suggests the need for learners to develop conscious understanding and knowledge of pronunciation features and deliberate actions to achieve them in their own speech—that is, for **metacognitive awareness** and control of their own pronunciation performance. Such a metacognitive orientation to pronunciation can be adopted when teaching both segmental and higher level pronunciation features, such as intonation and contrastive stress, as in Cauldwell's (2003, 2012) approach in his pronunciation teaching materials. Reed and Michaud (2015) suggest that teachers cannot assume learners have sufficiently learned such prosodic features simply by being able to repeat them, but need to also gain explicit awareness and understanding of the pragmatic functions of intonation in signaling speaker intent.

In sum, while implicit learning may help to imprint pronunciation form and patterning, it seems that techniques for encouraging noticing, attention, and conscious analysis and understanding of pronunciation forms and patterns and how these function in one's own performance and in communication with others are especially beneficial or necessary in L2 learning.

Teaching and Learning Activities and Techniques

Pronunciation Teaching and Learning Preliminaries

Learning the pronunciation of another language is a complex task which requires motivation, time, and patience on the part of the learner and teacher. Initially it is important for the teacher to spend some time building an awareness and concern for pronunciation, ensuring that learners (i) understand the component parts of pronunciation and how they work together and (ii) recognize the broader importance of pronunciation for communicative success in terms of intelligibility, fluency, impact, and presentation of self and identity. Pronunciation, more than any other area of language, is closely associated with emotion and identity and requires considerable sensitivity in terms of teaching activities and techniques. It is important for the teacher to address the emotional and attitudinal aspects of pronunciation learning in order to establish clear achievable goals and select activities and techniques appropriate to the learners.

Survey research suggests that the most common techniques used in pronunciation teaching include spontaneous error correction, reading aloud, phonetics training, and listen and repeat (Henderson et al., 2012; Tergujeff, 2013). Similar surveys suggest that the proportion of pronunciation-based activities in general language textbooks varies widely, with little or no coverage in some and extensive, systematic coverage in others. As a result, pronunciation teaching can be what Derwing and Munro (2015) refer to as “somewhat hit and miss” (p. 78). One way of avoiding this kind of hit-and-miss pronunciation teaching is by showing from the outset that there is more to pronunciation than learning individual sounds, that pronunciation is systematic, with the component parts interacting as a system in which syllable structure and prosody organizes individual consonants and vowels into larger rhythmic units and meaningful patterns. For instance, the teacher can show that the syllable, with a vowel as its central element, is fundamental to the construction of words and is the basic unit of rhythm in English. Explaining rhythmic patterning and the alternation of stressed and unstressed syllables should

help learners understand concepts such as vowel reduction, word stress, and prominence more easily.

Another way teachers can help learners understand that pronunciation learning is a systematic process is by teaching the different subskills learners need to acquire in order to master the various elements of pronunciation, including:

- (a) *noticing* – pronunciation elements in speech, similarities and differences between L1 and L2 pronunciation, both segmental phonemes and prosodies;
- (b) *discriminating* – between L1 and L2 elements, between correct and incorrect elements;
- (c) *imitating* – sounds and other elements of pronunciation accurately;
- (d) *reproducing* – elements without prompting;
- (e) *contextualizing* – individual elements within a stream of speech;
- (f) *generating* – pronunciations in new contexts;
- (g) *correcting* – their own inaccurate sounds and patterns.

(Rogerson-Revell, 2011, p. 212)

Learners and teachers need to be aware that developing pronunciation skills, from individual sounds to discourse-level intonation patterns, is a gradual process of acquisition involving all of these subskills, rather than just correcting the odd individual pronunciation error in an isolated listen-and-repeat session. The ultimate aim is for learners to be able to recognize and correct their own errors rather than rely on the teacher to do so, thus developing learner autonomy. Encouraging self, peer, and group correction from an early stage can facilitate this level of awareness.

Specific learning strategies for pronunciation can be identified as “deliberate actions and thoughts that are consciously employed, often in a logical sequence, for learning and gaining greater control over the use of various aspects of pronunciation” (Pawlak, 2010, p. 191). Following the categories of Oxford’s (1990) learning styles inventory, pronunciation learning strategies might include **cognitive strategies** such as listening, imitation, or repetition, as well as **metacognitive strategies** such as consulting a pronouncing dictionary, asking for pronunciation input

during interaction, or monitoring one's own performance for pronunciation errors and attempting self-correction. **Memory strategies** such as learning poems, songs, or speeches by heart may also help focus attention on pronunciation, as may **social strategies** such as seeking to associate and communicate with people whose pronunciation is considered to be attractive or a good model. In what follows, we review a range of strategies and approaches that can be recommended for pronunciation learning and teaching.

Critical Listening

A pronunciation learning strategy for raising metacognitive awareness that has been researched and is recommended by Fraser (2009) and by Couper (2011, 2015) for establishing category boundaries between phonemes, adjusting figure and ground in perception, and for eliminating unwanted sounds, is **critical listening**. This is a metacognitive strategy involving intensive listening specifically focused on the contrast between an acceptable or correct production and one that is unacceptable or incorrect, ideally contrasting “the learner’s production when it is acceptable with when it is not... in conjunction with getting feedback from peers or the teacher” (Couper, 2015, p. 426). Fraser (2009), for example, proposed the use of critical listening to help learners establish the boundaries between the /r/ and /l/ phonemes in English by focusing “on the contrast between a correct (or appropriate) pronunciation versus an incorrect (or inappropriate) pronunciation within a particular communicative act” (p. 301), while Couper (2011) demonstrated the value of critical listening for helping students to produce final voiced consonants without a following vowel sound, or **epenthetic vowel**.

Critical listening can be paired with other strategies to help students focus on sound and notice the details of pronunciation in authentic communicative contexts, as a way to retune their perception away from sound categories based in L1 pronunciation (L1 transfer) and the written language (sound-spelling correspondences). Strategies that can reinforce critical listening might include computer enhancements of speech or visuals of acoustic properties or articulatory movements, now widely

available in commercial programs and free online pronunciation tutorials (as reviewed in Chap. 5); shadowing (see below); impressionistic, sound-based spelling or drawing of intonation lines; or IPA transcription, which sharpens perception of how real language sounds and which students have been shown to value (Mompean & Lintunen, 2015). A contrastive approach can be recommended to sharpen perception of differences, such as focusing on the pronunciation of words spoken in isolation versus in phrases, or with exaggerated, hyperarticulation versus casual, conversational articulation; different realizations of a grammatical function (e.g., the intonation of *wh*-questions) or a pragmatic function (e.g., use of intonation in expressing politeness/impoliteness); differences between one accent and another; or divergences of a student's attempt at imitation from a pronunciation model.

Corrective Feedback

Most teachers and learners see feedback on errors as an important aspect of scaffolding language learning, though learners may sometimes feel that teachers do not correct them enough (Plonsky & Mills, 2006), and research shows that pronunciation teaching which includes feedback is more effective than instruction which does not (Lee et al., 2015). **Corrective feedback** (CF) has been defined as “responses to learner utterances containing an error” (R. Ellis, 2006, p. 28) and is the most frequent way that teachers engage with pronunciation in the classroom (Foote et al., 2013). Studies suggest that CF in combination with other kinds of FFI is an effective approach to teaching segmental aspects of pronunciation and that explicit feedback is more beneficial in raising awareness of errors than implicit feedback (Lyster, Saito, & Sato, 2013). Saito and Lyster (2012), for example, in their study of Japanese speakers learning English /r/, found that the most beneficial approach involved a combination of FFI and corrective feedback.

Long's (1996) **Interaction Hypothesis** suggests that corrective feedback during negotiation of meaning in communicative interaction might help to fine-tune implicit learning, and Swain's (1995) **Output Hypothesis** suggests that giving corrective feedback, as a form of input

that is explicitly geared to modifying a learner's output, could affect their learning by causing them to modify that output. Long (2015) reserves a special place for CF or other types of form-focused feedback given in response to a learner's specific act of communication, the moment an error is made, when the learner is likely to pay most attention to the feedback and also benefit from it the most, as a type of explicit instruction that can aid the implicit learning processes triggered in meaning-oriented tasks. Couper (2015) points out that whatever the type of CF, the fundamental concern is that the learner understands the correction. Beyond this immediate concern, a longer term concern is that CF (or any type of FFI) becomes implicit knowledge (R. Ellis, 2009, p. 18) and so leads to improved pronunciation outside the context in which the feedback was given, ultimately affecting learner language outside the classroom.

Physicality and Proprioception

Pronunciation, more than other areas of language learning, such as grammar and vocabulary, is physical. It involves the development of a set of muscular and mechanical habits in the L1 which have to be realigned when learning to speak another language. Underhill (2010) claims that the physicality of pronunciation means that “[i]t does not fit into a cognitive teaching paradigm” and makes it more like teaching dance than teaching grammar, though we note that from the perspective of cognitive phonology, pronunciation patterns exist at a cognitive level as well as a muscular and mechanical level (see Chap. 2), and these patterns can be said to have their own “grammar.” Underhill (2012) advises that teachers need to help learners become aware of the muscles that make sounds in order to develop their “proprioceptive intelligence” and retrain muscular habits. Developing this awareness is what is termed *proprioception*, a neurological concept referring to the internal sense of knowing which muscles and body parts one is using and with how much energy. It is worth noting that, as in the metacognitive view of pronunciation of cognitive phonology, this new “kinaesthetic” view of breaking the “pronunciation habit” of the mother tongue (Pennington, 1998) differs from notions of habit formation on which Audiolingualism was based in assuming that

speakers can develop conscious awareness and deliberate control of articulation.

Underhill (2010) claims that “most teachers do not know what is going on inside their and their students’ mouths,” and that if they are not aware of how sounds are physically articulated, they are largely restricted to listen-and-repeat or discrimination activities. He suggests that “... these are fine as far as they go, but they do not specifically set out to develop the proprioceptive or kinaesthetic intelligence that can gradually liberate the learner from the *oral* and *aural* grip of their mother tongue pronunciation habits” (Underhill, 2012, emphasis in original). Couper (2015), similarly, notes that teachers often do not have an accurate sense of the specific pronunciation features of the language they are teaching and how they are produced, so that both teachers and learners need to learn this.

Articulatory Setting

Underhill’s (2010, 2012) approach parallels to some extent earlier work incorporating teaching based on articulatory setting. Articulatory setting refers to the characteristic or long-term positioning of articulators by individuals or groups of speakers of a particular language or variety which create a particular voice quality or accent, that is, “the general differences in tension, in tongue shape, in pressure of the articulators, in lip and cheek and jaw posture and movement, which run through the whole articulatory process” (O’Connor, 1973, p. 289). It is this concept which underpins intuitive remarks about the characteristics of how a language sounds or how a speaker talks, such as “the English mumble, they don’t open their mouths when then talk,” or “German sounds quite harsh,” or “that man/woman’s got a really husky voice” (Rogerson-Revell, 2011, p. 37).

Although sometimes considered outside the domain of phonology per se, the importance of acquiring an L2 articulatory setting in pronunciation learning has been recognized for a long time (Honikman, 1964; Laver, 1978; O’Connor, 1973). Indeed, Honikman (1964) asserted that “where two languages are disparate in articulatory setting, it is not possible

to master the pronunciation of one whilst maintaining the articulatory setting of the other” (p. 298). Voice quality or accent resulting from articulatory setting could make a good starting point for working on pronunciation since a person’s first experience of an unknown language is usually an auditory impression of how it sounds, and many people are able to describe or imitate the way a language sounds before they actually know the language. Brown (1991) claimed that if students can be made aware of L2 articulatory settings from an early stage, if they can “get into gear” as Honikman (1964) described it, the large-scale adjustments in articulatory setting that produce a distinctive voice quality or accent will help them with the small-scale changes needed to articulate individual vowels and consonants, which Pennington (1989) maintained as a key rationale for a top-down approach to teaching pronunciation. Similarly, Jones and Evans (1995) claimed that “it gives students a chance to experience pronunciation on intuitive and communicative levels before moving on to a more analytical exploration of specific elements of phonology [and]...work in voice quality can help students to improve their image when they speak English, and thus increase their confidence” (pp. 245–246).

Such approaches are consistent with current language learning theory in their emphasis on developing noticing and awareness of language forms as well as with Underhill’s (2010, 2012) emphasis on the centrality of physicality and the need to “choreograph” new pronunciation patterns and “loosen the grip” of L1 phonological patterns. This physical approach can be seen as reigniting the idea of describing the articulation of sounds in order to get learners to produce them and also reemphasizing the importance of imitation and repetition to build new habits. Many pronunciation textbooks and software/apps provide visual representations of cross-sections of the vocal articulators and their movements in articulating sounds, and many teachers seek to describe these technically to students and then have them try to imitate and practice the unaccustomed movements. Using a combined kinaesthetic and metacognitive approach focused on articulatory setting (see Mompean, 2003, for pedagogical suggestions) reinforces the techniques that many teachers develop intuitively to help learners physically feel and reproduce target pronunciation features, whether individual sounds or word stress, without referring to

technical language to do so. When combined with perceptual training such as HVPT and/or FFI, such techniques seem promising for helping older learners adjust their L1-influenced pronunciation patterns as these exist on cognitive and motor levels in ways that interlink perception and production.

Shadowing

A technique that is increasingly used for pronunciation practice is “shadowing,” when the listener immediately tries to repeat what another speaker says. The technique is commonly used in the training of simultaneous interpreters (Lambert, 1992) but has also proved to be useful for language learning. Several studies have reported the effectiveness of shadowing for improving listening, reading, and speaking skills (Chung, 2010; Kusumoto, 2015; Shiki, 2011; Tamai, 2005). There have been fewer studies to date of the benefits for pronunciation, though Okada’s (2002) study of the use of shadowing to teach prosody suggests that shadowing helps improve rhythm.³ Similarly, in his study of Japanese learners of English, Kusumoto (2015) concluded that the use of shadowing improved their use of prosody. A recent study by Foote and McDonough (2017) in which participants practiced shadowing using short dialogues recorded on iPods demonstrated improvement after eight weeks in both shadowing and extemporaneous speaking measures of comprehensibility and fluency, though not accentedness.

As a strategy, shadowing can be considered to stand between cognitive and metacognitive form-focused techniques, in the sense that the learner performs the shadowing too quickly—in order to be able to keep up with the flow of speech in real time—for input to be fully analyzed and for output to be fully controlled. It stands out from other kinds of FFI that are widely used in pronunciation teaching and research in the combined broad scope and continuous high degree of attention required and also in the lesser extent of controlled processing involved. Different from critical listening, shadowing forces the learner to focus on form broadly or non-specifically and continuously, and at the same time to shortcut aspects of analysis which are not possible under conditions of hearing and repeating

speech as produced in ongoing talk. Given issues of cognitive load, shadowing as carried out by a language learner necessarily bypasses full language processing in favor of a kind of “light” or surface-oriented processing of heard speech which may help develop new cognitive representations for fluent speech that are more accurately based in sound.

Bridging Activities

Pronunciation activities can take many forms, from very controlled practice of individual phonemes to freer, communicative activities focusing primarily on fluency and meaning. Muller Levis and Levis (2016) suggest that a category of activities is needed between these controlled and communicative activities, which they refer to as “bridging activities,” in which the learners focus on both form and meaning and in which pronunciation is key to communicative success. They advise teaching prominence alongside the teaching of greetings and small talk, within communicative chunks, as in the following exchange:

Monika: HelLO Jim how ARE you

Jim: I'm FINE How are YOU
(adapted from Muller Levis & Levis, 2016, p. 232)

This approach incorporates the notion of teaching learners chunks of language rather than individual lexical items. The idea that language teachers should take account of conventionalized strings of words or lexical chunks is not new (see, e.g., Palmer, 1924) but was given little priority until relatively recently, after increased attention to this feature of language in SLA studies (see Ellis & Wulff, 2015, for historical overview), and has been reinforced by corpus-based research on spoken language. McCarthy and O'Dell's (2005) work on North American English conversation from the *Cambridge International Corpus* (CIC) explored the prevalence of lexical chunks such as *I don't know if* and *a lot of people* in everyday spoken language. Learners' use of such chunks has been shown to correlate positively with L2 proficiency and fluency (Lindstromberg & Boers, 2008) and has been recommended for pronunciation teaching,

particularly for intonation practice (Reed & Michaud, 2015; Seidlhofer & Dalton-Puffer, 1995).

Data-Driven Learning

Linguists have been using computer corpora (a **corpus** is a body of authentic language data, and **corpora** is the plural form) for research purposes since the 1980s. The potential of corpora for language learning and teaching was seen fairly quickly, for instance, with the COBUILD (Collins Birmingham University International Language Database) dictionary project in 1987 which led to the use of the corpus in the development of the COBUILD language teaching resources (Collins, 2011). The term **data-driven learning** (DDL) emerged at this time to describe how learners could learn about language by exploring language data themselves (Johns, 1986). Since then, there has been growing interest in the pedagogic use of corpora for language learning (Aston, 1995; Gavioli, 2001; Sinclair, 2004), particularly for teaching grammar (e.g. Vannestål & Lindquist, 2007) and vocabulary (Yilmaz & Soruč, 2015).

Speech corpora have played a similarly important role in phonetic and phonological research, but their application to pronunciation teaching and learning is still relatively limited. There are however many corpora of spoken English which offer opportunities for teachers and learners to collect and analyze speech samples for different varieties of English and in different types of speech functions and fields. The Speech Accent Archive,⁴ for example, contains over 2000 samples of L1 and L2 English speakers, with participants from 246 native language backgrounds. All samples have audio files and many also have detailed phonetic transcriptions. The Santa Barbara Corpus of Spoken American English⁵ also includes transcriptions with varying levels of phonetic detail. The VOICE (Vienna-Oxford International Corpus of English) corpus⁶ provides a database of spoken interactions by 1250 speakers, with approximately 50 different first languages, in a range of ELF contexts.

Little research has been conducted to date on the effectiveness of data-driven approaches to pronunciation teaching and learning. However, there is some indication of beneficial effects. For instance, Gut (2005) found that using a corpus-based approach as part of a pronunciation

training course in German resulted in significant improvements in participants' prosodic knowledge and in correct stress placement. Cao (2016) described the development and use of an English speech corpus of Mandarin learners to enable teachers and learners to compare the speech patterns of native English and Mandarin speakers of English.

As with the use of technology generally for language learning, consideration needs to be given to how best to incorporate such approaches into classroom teaching or self-study (see Chap. 5 for more on the issue of technology versus pedagogy). Data-driven approaches generally lend themselves well to awareness-raising and problem-solving tasks, such as focusing on a particular phonetic feature (e.g., the occurrence of schwa or contracted forms) in different speech samples, or project work in which students are tasked with making their own corpus by collecting a set of speech samples in one or more speech communities.

Pronunciation Learning Strategies of High Performance Learners

In decisions about what strategies to use for working on pronunciation, teachers might look to those used by learners judged to have high performance in pronunciation. Eckstein's (2007) study of the pronunciation learning strategies and speaking skills of international student learners of English as L2 in the United States found three strategies related to high pronunciation performance: strategies for noticing pronunciation mistakes, adjusting facial muscles while speaking, and asking for help with the pronunciation of new words. The first of these can be considered a form of critical listening, the second a strategy involving proprioceptive feedback, and the third a social strategy involving seeking out "expert" or native speaker input. It is noteworthy that all involve raised awareness and can thus be considered metacognitive strategies. Szyszka's (2015) study of two groups that she terms "good" and "average" "pronunciation users" employed a reduced version of a Pronunciation Learning Strategies Inventory (Berkil, 2009) based on the language learning styles inventory of Oxford (1990) and also asked open-ended questions about participants' preferred pronunciation learning strategies. Szyszka's (2015, p. 99)

findings can be summarized in terms of the most frequently to least frequently mentioned strategies (the strategies in rank 7 were all mentioned by just one person):

Cognitive

- 1 – Listening (to native speakers who provided a good model and to songs and movies)
- 2 – Imitation (mimicking native speaking interlocutors, actors/actresses, and pop singers)
- 3 – Repetition (of speech samples)
- 4 – Talking (to oneself and to native speakers)
- 5 – Singing songs and reading texts out loud

Metacognitive

- 6 – Checking or studying pronunciation in a dictionary
- 7 – Teaching pronunciation, transcribing, following a notice-and-improve approach

Memory

- 7 – Learning songs by heart

Affective

- 7 – Feeling good about correct pronunciation.

The “good pronunciation users” further indicated that, in addition to concern for L2 pronunciation, they see their exposure to the L2 outside of class as affecting their pronunciation, along with the different strategies for learning pronunciation. We note that the learners with good pronunciation skills care about pronunciation and realize that they can learn pronunciation in social context, and Szyszka (2015) stresses these good pronunciation learners’ recognition of “the high value of a strategic approach to pronunciation learning” (p. 100).

Implications of Current Views of Language and Learning for the Teaching of Pronunciation

As comes out from the discussion in this chapter, in the context of the discussions in Chaps. 2 and 3, changing views of language and of learning have impacted the teaching of language and specifically pronunciation

over a long period. Here we discuss three current orientations to language and learning and their implications for teaching.

Socially Oriented Approaches to Pronunciation Learning in Contexts of Multilingualism, Plurilingualism, and Translanguaging

The first of these is the new focus on sociolinguistically informed, multi-varietal and multilingual approaches to language learning in contexts of multilingualism, plurilingualism, and translanguaging (Chap. 3) aiming to develop a broad repertoire of pronunciation features and competences. An important implication is that a range of accents and speech styles be presented as a basis for the learner to develop a rich internal model of the phonological system of the L2 and a rich and diverse set of pronunciation features to draw on in performance. These would obviously have to be selected or adapted to correspond to the proficiency level and needs of the learner. Some examples include:

- Developing a list of L1 loanwords from the L2 and practicing the pronunciation of the two sets of words with the goals of developing increased awareness of how differently they are pronounced in the two languages, being able to differentiate them at will, and automatizing the pronunciation of the L2 words as desired (for further ideas, see Rogerson-Revell, 2011, p. 214);
- Listening to, imitating, and discussing impressions of a wide range of accents;
- Imitating a high-profile speaker from a recording or video available on TV or the internet and adjusting performance until listeners (e.g., classmates or friends) rate the match as good;
- Pronouncing English in a rhotic vs. a non-rhotic way;
- Speaking English with a British, American, Australian, or Indian accent;
- Imitating very careful, form-focused pronunciation vs. very relaxed and casual pronunciation;
- Varying common set phrases such as greetings and leave takings, formulas for complimenting and apologizing;

- Altering prosody in constructed tasks in which a speaker has instructions to diverge or converge with another speaker (e.g., angry vs. conciliatory, excited vs. calm, accepting vs. questioning, highly engaged vs. disengaged with the topic); and
- Roleplaying to practice style-shifting and projecting different identities associated with different features of pronunciation and accent.

In this way, students will be developing flexibility in their pronunciation as they build their sociolinguistic and multilingual repertoire of pronunciation features and phonologically based meaning potential for their own speech.

Social Learning Theories and Cognitive Linguistics

Couper (2015) offers a discussion of language and learning theory and its relation to pronunciation practice that is a valuable updating and theoretical grounding of the field. He considers that the “social turn” in second language acquisition (SLA; see Block, 2003) makes theories of learning from education and psychology as well as social theories of potential relevance for teaching. In his view, Cognitive Linguistics “provides a useful framework for bringing together both the cognitive and social aspects of pronunciation learning” (Couper, 2015, p. 414). Based on his review of social theories of learning from Educational Psychology, together with SLA theory and research, L2 phonology and speech, and Cognitive Linguistics and phonology, Couper (2015) seeks to translate insights from these theoretical strands into pronunciation teaching practice. To this end, he gives five tips for teachers, suggesting “classroom activities only need to be adjusted slightly” (p. 429) to implement the teaching approaches he advocates, which are broadly in line with Cognitive Phonology.

1. *Teaching tip one: understand all is not as it seems*: “move away from your subjective perception of L1 speech” (p. 422), such as the literate bias (Linell, 2005/1982) that leads speakers to perceive speech as made up of invariant phonemes. As Couper (2015) elaborates, “the main point is to learn to be able to step aside from your own perceptions...and listen to

sounds more neutrally to try and imagine how they might sound to your students” (p. 424). Students will need to learn to notice cues for L2 phonemic differences that are different from those they use for recognizing L1 phonemic differences, and to do so, they will need to pay attention to how speech is actually pronounced. Teachers therefore need to be aware of their own biases tied to their phonemic and literate perception as a result of their long history of speaking and literacy with their native language and work on learning to hear without these filters. In our view, learning the International Phonetic Alphabet (IPA) is a good first step.

2. *Teaching tip two: generate dialogue.* Couper suggests that teachers need to develop a “socially constructed metalanguage” (p. 424) interactively with students for talking about pronunciation. For example, in Couper (2013, p. 9), he suggests using learners’ own descriptions of how to correct syllable codas, such as “that’s too strong,” “say it shorter,” “it becomes quiet,” or “make it smaller.” Couper cites research he has done showing the value of a teacher dialoguing with students about the teacher’s and the students’ perceptions of the L2 and of their own speech.

3. *Teaching tip three: establish category boundaries through Critical Listening.* As Couper (2015), drawing on the work of Fraser (2009), describes it:

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.... 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. (p. 426)

Couper suggests recording students at the beginning of a course and giving feedback on their pronunciation that they can use “as a diagnostic” (p. 426) as they “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” (p. 427).

4. *Teaching tip four: meaningfully integrate pronunciation into further practice activities.* Couper (2015) gives examples from his own prior work to suggest ways in which “teachers can draw on their experience with communicative language teaching to devise activities that will support the development of phonological concepts...” (p. 427). Many of the activities he suggests involve a focus on form within a communicative task or as preparation or reflective follow-up to communicative tasks.

5. *Teaching tip five: provide the right kind of corrective feedback.* As Couper reviews from his previous work, students do not always get the intended benefit from corrective feedback:

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. Couper (2015, p. 428)

A teaching sequence that is broadly consistent with what Couper (2015) advocates is that of Celce-Murcia et al. (2010, pp. 44–49), which starts with a focus on form and then increasingly focuses on meaning and communication while maintaining a focus on form through corrective feedback in communicative tasks. Its steps are: (1) Description and analysis, (2) Listening discrimination, (3) Controlled practice, (4) Guided practice, and (5) Communicative practice. This is similar to the curriculum proposed by Pennington (1996, pp. 224–228), which starts with focusing attention on form and progresses to interactive communication:

Unit structure	Activity type	Practice level	Cognitive load	Modality	Participation	Information
Presentation	Focus	Mechanical	Low inference	- Production	- Interaction	- Communicative
	Contextualization	Contextualized				
Practice	Controlled	Meaningful				
	Structured	Realistic				
	Free	Real	High inference	+ Production	+ Interaction	+ Communicative

(Pennington, 1996, p. 226, Fig. 6.2)

Celce-Murcia et al. (2010) and Pennington (1996), like Couper (2015), based their ideas for teaching sequences on theory and research findings.

Pronunciation Technologies

The final orientation is to computer-based, internet-delivered, and wireless technologies for pronunciation instruction. Use of pronunciation technologies is based on theories of learning suggesting that audio-visual and technological aids can promote learning by enhancing motivation, memorability, and engagement through offering different and potentially more effective and efficient learning modalities than those provided in traditional teaching–learning situations and human interactions. With access to the internet, the capabilities of pronunciation technology for learning, both in class and out of class, are broadened to include access to a wide range of resources of other people, information sources, and technological aids—all of which potentially enhance the learning experience and increase the learnability of pronunciation. When married to the portability of smartphone technology, the ability to learn “on demand” makes any time a possible learning occasion and any place a possible learning environment. In its various types of modalities, pronunciation technology offers:

- improved quality and variety of analysis and feedback on speech;
- new types of learning experiences;
- opportunities for individualized study according to a learner’s own needs;
- increased quantity, quality, and variety of access to language; and
- new types of access to speakers of other languages.

The technologies available for pronunciation in individual packages and through internet access provide computer-aided pronunciation (CAP) input that is quick, repeatable, precise, reliable, authoritative, highly salient, multi-modal, individual, and variable (Pennington, 1999, p. 430). Drawing on the “pros” of CAP pedagogy delineated by

Pennington (1999, p. 43), we suggest that in the best cases such pedagogy is:

- motivating;
- stimulates effort;
- raises awareness of pronunciation;
- increases understanding of the nature of pronunciation;
- enhances learnability of distinctions and patterns which may otherwise not be attended to;
- increases automaticity of performance by providing opportunities for focused practice;
- fosters precision in articulating distinctions;
- builds confidence based on greater access to information and learning aids; and
- develops pronunciation skills.

Since use of technology has become a very big part of the teaching of pronunciation, we have given it a chapter of its own (Chap. 5).

Concluding Remarks

Many language learners struggle with pronunciation even when they have achieved a reasonable level of mastery of grammar and vocabulary. However, research has shown that despite the importance often placed on pronunciation by students and teachers, there is still relatively little emphasis on pronunciation in the language classroom (Derwing & Rossiter, 2002; Henderson et al., 2012). This may partially reflect limited teaching time and the pressure of syllabuses and exams which often do not give priority to pronunciation, regardless of the real-world value of being able to speak intelligibly and fluently. Pronunciation needs to be seen as fundamental for successful spoken communication. It is important for teachers to have an awareness and concern for pronunciation and to convey this to their students. Learning the pronunciation of a language is a complex and often daunting task; but as with other areas of language pedagogy, there are approaches and techniques that can be used

to facilitate the process. Having a critical awareness of pedagogic resources and related research can help teachers make informed choices about what and how to teach.

As with any area of language teaching, a key concern is how effective instruction is; and although research focusing on this question is relatively recent, it is showing that pronunciation teaching can have a positive effect. Yet much more remains to be done to provide teachers, curriculum planners, and materials developers with the evidence they need to make informed choices about what and how to teach pronunciation. As we have said, this reflects the ongoing gap between research and practice in the field. On the one hand, many techniques and methods that are receiving significant attention from teachers and other practitioners may nonetheless not yet be sufficiently grounded in research for applied linguists, teacher educators, and other specialists to have confidence in recommending their use. This is in part because, as we observed above with respect to the technique of shadowing, there is not yet a sufficiently established research base in international refereed journals (though the recent study by Foote & McDonough, 2017, which was published in the *Journal of Second Language Pronunciation*, represents an encouraging trend in terms of positive research findings for the shadowing technique using technology-based teaching methodology). On the other hand, even though there are very active areas of applied pronunciation research, such as on FFI teaching techniques, those techniques are not widely applied. One reason is that teachers may not be reading the journals in which the research studies are appearing. This presents a paradoxical situation, since publication in a research-oriented journal, especially one with international scope that publishes relatively technical articles, is a key indicator of high quality research, and yet the authority gained from such publication makes it less accessible to teachers. Another reason is that research studies do not generally give sufficient information about how to apply a particular method for readers to be able to adopt it in their own teaching contexts.

In the chapters to come, we will continue reviewing the current state of pronunciation practice and the existing resources and approaches that practitioners can draw on, as well as reviewing the available evidence relating to specific practices and pointing out gaps in that evidence that

indicate future research directions and needs. In this way, we hope to contribute to the synergy between research and practice that ensures ongoing development and progress on both sides.

Notes

1. This can be seen as the antithesis to the form-focused approaches described by Wulff and Ellis (2018) which aim to produce an L2-tuned linguistic system by providing micro-level form-focused instruction first, with a goal of noticing and conscious processing through explicit learning before subsequent implicit processing takes place that will consolidate and systematize what has been noticed and consciously learned.
2. Input enhancement might potentially include the “hyperarticulation” (Lindblom, 1990) which is characteristic of sounds and words spoken in isolation and various types of auditory and visual enhancement by digital means (see Fouz-González, 2015, for examples).
3. It should be noted that although shadowing is attracting a lot of attention in language teaching and research (notably in Japan), a number of the studies on shadowing have not appeared in international refereed journals and so the database for this technique as applied to language learning is not as strong as in some other areas of pronunciation practice.
4. <http://accent.gmu.edu/>
5. <http://www.linguistics.ucsb.edu/research/santa-barbara-corpus>
6. <https://www.univie.ac.at/voice/page/index.php>

References

- Altinsel, Z., & Rittenberg, W. (1996). *Cultural support for international TAs: An undergraduate buddy program*. Paper presented at the Conference of Teachers of English to Speakers of Other Languages, Chicago, IL, March 26–30, 1996. ERIC Document No. ED407918.
- Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speakers’ judgements of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning*, 42(4), 529–555. <https://doi.org/10.1111/j.1467-1770.1992.tb01043.x>

- Aston, G. (1995). Corpora in language pedagogy: Matching theory and practice. In G. Cook & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics: Studies in honour of H.G. Widdowson* (pp. 257–270). Oxford: Oxford University Press.
- Bailey, K. M. (1983). Foreign teaching assistants at U.S. universities: Problems in interaction and communication. *TESOL Quarterly*, 17(2), 308–310. <https://doi.org/10.2307/3586658>
- Bailey, K. M., Pialorsi, F., & Zukowski-Faust, J. (Eds.). (1984). *Teaching assistants in U. S. Universities*. Washington, DC: National Association for Foreign Student Affairs.
- Beinhoff, B. (2013). *Perceiving identity through accent: Attitudes towards non-native speakers and their accents in English*. Oxford and Bern: Peter Lang.
- Berkil, G. (2009). *A closer look at pronunciation learning strategies*. Saarbrücken: VDM—Publishing Cited in Szyszka (2015).
- Block, D. (2003). *The social turn in second language acquisition*. Edinburgh: Edinburgh University Press.
- Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics*, 61(5), 977–985. <https://doi.org/10.3758/BF03206911>
- Brazil, D., Coulthard, M., & Johns, C. (1980). *Discourse intonation and language teaching*. Harlow, UK: Longman.
- Bresnahan, M. J., Ohashi, R., Nebashi, R., Liu, W. Y., & Shearman, S. M. (2002). Attitudinal and affective response toward accented English. *Language & Communication*, 22(2), 171–185. [https://doi.org/10.1016/S0271-5309\(01\)00025-8](https://doi.org/10.1016/S0271-5309(01)00025-8)
- Brown, A. (1991). *Pronunciation models*. Singapore: Singapore University Press.
- Brumfit, C. J., & Johnson, K. (1979). *The communicative approach to language teaching*. Oxford: Oxford University Press.
- Byrd, P., Constantinides, J. C., & Pennington, M. C. (1989). *Foreign teaching assistant's manual*. New York: Collier Macmillan.
- Canagarajah, A. S. (1999). Interrogating the “native speaker fallacy”: Non-linguistic roots, non-pedagogical results. In G. Braine (Ed.), *Non-native educators in English language teaching* (pp. 77–92). Mahwah, NJ: Erlbaum.
- Cao, R. (2016). Improving English pronunciation teaching and learning via speech corpora of learners with dialectal backgrounds. *IJet*, 11(4), 90–94. <https://doi.org/10.3991/ijet.v11i04.5461>

- Cauldwell, R. (2003). *Streaming speech: Listening and pronunciation for advanced learners of English*. [Student's book]. Birmingham: Speechinaction. ISBN 0-9543447-1-5.
- Cauldwell, R. (2012). *Cool speech app*. Retrieved July 30, 2017, from <http://www.speechinaction.org/cool-speech-2/>
- Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M., with Griner, B. (2010). *Teaching pronunciation: A course book and reference guide* (2nd ed.). New York: Cambridge University Press.
- Chambers, D. (2015). *Working with teaching assistants and other support staff for inclusive education*. Bingley, UK: Emerald.
- Chung, D.-U. (2010). The effect of shadowing on English listening and speaking abilities of Korean middle school students. *English Teaching*, 65(3), 97–127. <https://doi.org/10.15858/engtea.65.3.201009.97>
- Cintrón-Valentín, M., & Ellis, N. C. (2016). Salience in second language acquisition: Physical form, learner attention, and instructional focus. *Frontiers in Psychology*, 7. Language Sciences, Special Topic: Perceptual linguistic salience: Modeling causes and consequences. <https://doi.org/10.3389/fpsyg.2016.01284>
- Collins. (2011). COBUILD reference resources. Retrieved February 26, 2018, from <https://collins.co.uk/category/English+Language+Teaching/COBUILD+Reference/>
- 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. <https://doi.org/10.1080/09658416.2011.570347>
- Couper, G. (2013). Talking about pronunciation: Socially constructing meta-language. *English Australia*, 29(1), 3–18. Retrieved August 12, 2017, from <http://ejournal.realviewdigital.com/?iid=111168#folio=1>
- Couper, G. (2015). Applying theories of language and learning to teaching pronunciation. In M. Reed & J. M. Levis (Eds.), *Handbook of English pronunciation* (pp. 413–432). New York: John Wiley & Sons.
- Dalton, C., & Seidlhofer, B. (1994). *Pronunciation*. Oxford: Oxford University Press.
- de Jong, N. H., Steinel, M. P., Florijn, A. F., Schoonen, R., & Hulstijn, J. H. (2012). Facets of speaking proficiency. *Studies in Second Language Acquisition*, 34(1), 5–34. <https://doi.org/10.1017/S0272263111000489>
- Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition*, 19(1), 1–16. Retrieved January 10, 2018, from <https://www.cambridge.org/core/journals/studies-in-second-language-acquisition/issue/77E9B64BB6D1D3DC15A6523E69A42665>

- Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly*, 39(3), 379–397. <https://doi.org/10.2307/3588486>
- Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. Amsterdam and Philadelphia: John Benjamins.
- Derwing, T. M., & Rossiter, M. J. (2002). ESL learners' perceptions of their pronunciation needs and strategies. *System*, 30(2), 155–166. [https://doi.org/10.1016/S0346-251X\(02\)00012-X](https://doi.org/10.1016/S0346-251X(02)00012-X)
- Doughty, C., & Williams, J. (1998). Pedagogical choices in focus on form. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 197–262). New York, NY: Cambridge University Press.
- Eckstein, G. T. (2007). *A correlation of pronunciation learning strategies with spontaneous English pronunciation of adult ESL learners*. MA thesis, Brigham Young University. Cited from Szyszka (2015).
- Ellis, N. C. (1994). Implicit and explicit processes in language acquisition: An introduction. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 1–32). London: Academic Press.
- Ellis, N. C. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition*, 18(1), 91–126. <https://doi.org/10.1017/S0272263100014698>
- Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24(2), 143–188. <https://doi.org/10.1017/S0272263102002024>
- Ellis, N. C. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition*, 27(2), 305–352. <https://doi.org/10.1017/S027226310505014X>
- Ellis, N. C., & Wulff, S. (2015). Second language acquisition. In E. Dabrowska & D. Divjak (Eds.), *Handbook of cognitive linguistics* (pp. 409–431). Berlin: DeGruyter Mouton.
- Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.
- Ellis, R. (2003). *Task-based language teaching and learning*. Oxford: Oxford University Press.
- Ellis, R. (2006). Researching the effects of form-focused instruction on L2 acquisition. *AILA Review*, 19, 18–41. <https://doi.org/10.1075/aila.19.04ell>
- Ellis, R. (2009). Implicit and explicit learning, knowledge and instruction. In R. Ellis, S. Loewen, C. Elder, H. Reinders, R. Erlam, & R. Philp (Eds.),

- Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 3–25). Bristol, UK: Multilingual Matters/ Channel View Publications.
- Esling, J. H. (1987). Methodology for voice setting awareness in language classes. *Revue de Phonétique Appliquée*, 85, 449–473.
- Field, J. (2003). Promoting perception: Lexical segmentation in L2 listening. *ELT Journal*, 57(4), 325–333. <https://doi.org/10.1093/elt/57.4.325>
- Foote, J. A., & McDonough, K. (2017). Using shadowing with mobile technology to improve L2 pronunciation. *Journal of Second Language Pronunciation*, 3(1), 34–56. <https://doi.org/10.1075/jslp.3.1.02foo>
- Foote, J. A., Trofimovich, P., Collins, L., & Urzúa, F. (2013). Pronunciation teaching practices in communicative second language classes. *Language Learning Journal*, 44(2), 181–196. <https://doi.org/10.1080/09571736.2013.784345>
- Fouz-González, J. (2015). Trends and directions in computer assisted pronunciation training. In J. Mompean & J. Fouz-González (Eds.), *Investigating English pronunciation: Trends and directions* (pp. 314–342). Basingstoke, UK and New York: Palgrave Macmillan.
- Fraser, H. (2009). Pronunciation as categorization: The role of contrast in teaching English /t/ and /l/. In A. Mahboob & C. Lipovsky (Eds.), *Studies in applied linguistics and language learning* (pp. 289–306). Newcastle upon Tyne: Cambridge Scholars Publishing.
- Gavioli, L. (2001). The learner as researcher: Introducing corpus concordancing in the classroom. In G. Aston (Ed.), *Learning with corpora* (pp. 108–137). Houston: Athelstan.
- Goh, C. C. M. (2000). A cognitive perspective on language learners' listening comprehension problems. *System*, 28(1), 55–75. [https://doi.org/10.1016/S0346-251X\(99\)00060-3](https://doi.org/10.1016/S0346-251X(99)00060-3)
- Goo, J., Granena, G., Yilmaz, Y., & Novella, M. (2015). Implicit and explicit instruction in L2 learning. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 443–482). Amsterdam: John Benjamins.
- Gorsuch, G. J. (2011). Improving speaking fluency for international teaching assistants by increasing input. *TESL-EJ*, 14(4), 1–25. Retrieved September 11, 2017, from <http://www.tesl-ej.org/wordpress/issues/volume14/ej56/ej56a1/>. Also available at https://www.researchgate.net/publication/313892087_Improving_Speaking_Fluency_for_International_Teaching_Assistants_by_Increasing_Input
- Gorsuch, G. (Ed.). (2015). *Talking matters: Research on talk and communication of international teaching assistants*. Stillwater, OK: New Forums Press.
- Gorsuch, G. (2016). International teaching assistants at universities: A research agenda. *Language Teaching*, 49(2), 275–290. <https://doi.org/10.1017/S0261444815000452>

- Gorsuch, G. (2017). *An evaluation of a teaching practicum course for international teaching assistants*. Working Paper, Texas Tech University. <https://doi.org/10.13140/RG.2.2.31122.45765>
- Gorsuch, G., Meyers, C. M., Pickering, L., & Griffiee, D. T. (2013). *English communication for international teaching assistants* (2nd ed.). Long Grove, IL: Waveland Press.
- Gut, U. (2005). Corpus-based pronunciation training. In *Proceedings of the Pronunciation Teaching and Learning Conference*, University College London. Retrieved February 26, 2018, from https://www.ucl.ac.uk/pals/study/cpd/cpd-courses/ptlc/proceedings_2005/ptlcp27
- Hahn, L. D. (2004). Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly*, 38(2), 201–223. <https://doi.org/10.2307/3588378>
- Hardison, D. M. (2004). Generalization of computer-assisted prosody training: Quantitative and qualitative findings. *Language Learning and Technology*, 8(1), 34–52. <http://dx.doi.org/10125/25228>
- Henderson, A., Frost, D., Tergujeff, E., Kautzsch, A., Murphy, D., Kirkovana-Naskova, A., et al. (2012). The English pronunciation teaching in Europe survey: Selected results. *Research in Language*, 10, 5–27. https://doi.org/10.1057/9781137509437_12
- Higgs, T., & Clifford, R. (1982). The push towards communication. In T. Higgs (Ed.), *Curriculum, competence, and the foreign language teacher* (pp. 57–79). Skokie, IL: National Textbook Company.
- Honikman, B. (1964). Articulatory settings. In D. Abercrombie, D. B. Fry, P. A. D. MacCarthy, N. C. Scott, & J. L. M. Trim (Eds.), *In honour of Daniel Jones* (pp. 73–84). London: Longman.
- Isaacs, T., & Trofimovich, P. (2012). Deconstructing comprehensibility. *Studies in Second Language Acquisition*, 34(3), 475–505. <https://doi.org/10.1017/S0272263112000150>
- 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(3), 535–543. <https://doi.org/10.2307/3588493>
- Johns, T. (1986). Micro-Concord: A language learner's research tool. *System*, 14(2), 151–162. [https://doi.org/10.1016/0346-251X\(86\)90004-7](https://doi.org/10.1016/0346-251X(86)90004-7)
- Jones, R. H., & Evans, S. (1995). Teaching pronunciation through voice quality. *English Language Teaching Journal*, 49(3), 244–251. <https://doi.org/10.1093/elt/49.3.244>

- Kachru, B. B. (1985). Standards, codification and sociolinguistic realism: The English language in the outer circle. In R. Quirk & H. Widdowson (Eds.), *English in the world: Teaching and learning the language and literatures* (pp. 11–30). Cambridge: Cambridge University Press.
- Kang, O., Rubin, D., & Pickering, L. (2010). Suprasegmental measures of accentedness and judgments of language learner proficiency in oral English. *The Modern Language Journal*, 94(4), 554–566. <https://doi.org/10.1111/j.1540-4781.2010.01091.x>
- Klein, W., & Dimroth, C. (2009). Untutored second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *The new handbook of second language acquisition* (pp. 503–521). Bingley, UK: Emerald.
- Koda, K. (2007). Phonology in literacy. In M. C. Pennington (Ed.), *Phonology in context* (pp. 219–244). Basingstoke, UK and New York: Palgrave Macmillan.
- Krashen, S. (1982). *Principles and practice in second language acquisition*. Oxford: Pergamon Press.
- Krashen, S. (1985). *The input hypothesis: Issue and implications*. New York: Longman.
- Krashen, S., & Terrell, T. (1983). *The natural approach: Language acquisition in the classroom*. Oxford: Pergamon.
- Kusumoto, Y. (2015). Exploring the effects of shadowing on prosody. Proceedings of the 4th International Conference on English Pronunciation (EPIP4): Issues & Practices conference, Prague, May 2015 (pp. 82–85).
- Lambacher, S. G., Martens, W. L., Kakehi, K., Marasinghe, C. A., & Molholt, G. (2005). The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Applied Psycholinguistics*, 26(2), 227–247. <https://doi.org/10.1017/S0142716405050150>
- Lambert, S. (1992). Shadowing. *Métra*, 37(2), 263–273. <https://doi.org/10.7202/003378ar>
- Laver, J. (1978). The concept of articulatory settings: An historical survey. *Historiographia Linguistica*, 5, 1–14. <https://doi.org/10.1075/hl.5.1-2.02lav>
- Leather, J. (1999). Second-language speech research: An introduction. *Language Learning*, 49(s1), 1–56. <https://doi.org/10.1111/0023-8333.49.s1.1>
- Lee, J., Jang, J., & Plonsky, L. (2015). The effectiveness of second language pronunciation instruction: A meta-analysis. *Applied Linguistics*, 36(3), 345–366. <https://doi.org/10.1093/applin/amu040>
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

- Levelt, W. J. M. (1999). Producing spoken language: A blueprint of the speaker. In C. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 83–122). Oxford, UK: Oxford University Press.
- Levis, J. M. (2005). Changing contexts and shifting paradigms. *TESOL Quarterly*, 39(3), 369–377. <https://doi.org/10.2307/3588485>
- Levis, J. (2007). Computer technology in teaching and researching pronunciation. *Annual Review of Applied Linguistics*, 27, 184–202. <https://doi.org/10.1017/S0267190508070098>
- Levis, J. M., & Grant, L. (2003). Integrating pronunciation into ESL/EFL classrooms. *TESOL Journal*, 12, 13–19. <https://doi.org/10.1002/j.1949-3533.2003.tb00125.x>
- Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and modeling* (pp. 403–439). Amsterdam: Kluwer Academic.
- Lindstromberg, S., & Boers, F. (2008). Phonemic repetition and the learning of lexical chunks. The power of assonance. *System*, 23(4), 423–436. <https://doi.org/10.1016/j.system.2008.01.002>
- Linell, P. (2005/1982). *The written language bias in linguistics: Its nature, origins, and transformations*. London and New York: Routledge. Earlier version without subtitle published 1982 in *Studies in Communication*, 2. Linköping, Sweden: Department of Communication Studies, University of Linköping. Retrieved from <https://www.diva-portal.org/smash/get/diva2:757560/FULLTEXT01.pdf>
- Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /ɹ/ and /l/: A first report. *Journal of the Acoustical Society of America*, 89(2), 874–886. <https://doi.org/10.1121/1.1894649>
- Long, M. H. (1981). Input, interaction and second language acquisition. In H. Winitz (Ed.), *Native and foreign language acquisition* (pp. 259–278). Annals of the New York Academy of Sciences, 379. New York: New York Academy of Sciences. <https://doi.org/10.1111/j.1749-6632.1981.tb42014.x>
- Long, M. H. (1983). Native speaker/non-native speaker conversation and the negotiation of comprehensible input. *Applied Linguistics*, 4(2), 126–141. <https://doi.org/10.1093/applin/4.2.126>
- Long, M. H. (1991). Focus on form: A design feature in language teaching methodology. In K. de Bot, R. B. Ginsberg, & C. Kramsch (Eds.), *Foreign language research in cross-cultural perspective* (pp. 39–52). Amsterdam: John Benjamins.

- Long, M. H. (1996). The role of linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). New York: Academic Press.
- Long, M. (2015). *Second language acquisition and task-based language teaching*. Malden, MA and Oxford, UK: Wiley Blackwell.
- Long, M. H., Gor, K., & Jackson, S. (2012). Linguistic correlates of second language proficiency. *Studies in Second Language Acquisition*, 34(1), 99–126. <https://doi.org/10.1017/S0272263111000519>
- Lyster, R., Saito, K., & Sato, M. (2013). Oral corrective feedback in second language classrooms. *Language Teaching*, 46(1), 1–40. <https://doi.org/10.1017/S0261444812000365>
- MacDonald, S. (2002). Pronunciation views and practices of reluctant teachers. *Prospect*, 17(3), 3–16. Retrieved August 15, 2017, from http://www.ameprc.mq.edu.au/docs/prospect_journal/volume_17_no_3/17_3_1_MacDonald.pdf
- McCarthy, M., & O'Dell, F. (2005). *English collocations in use: How words work together for fluent and natural English*, Intermediate. Cambridge, UK: Cambridge University Press.
- Mompean, J. A. (2003). Pedagogical tools for teaching articulatory setting. In M. J. Solé & D. Recansens (Eds.), *Proceedings of the 15th International congress of phonetic sciences* (pp. 1603–1606). Adelaide, Australia: Causal Productions.
- Mompean, J. A. (2006). Introduction: Cognitive phonology in cognitive linguistics. *International Journal of English Studies*, 6, vii–xii. Retrieved August 17, 2017, from <https://www.scribd.com/document/298249578/Cognitive-Phonology-Ijes-6-2>
- Mompean, J. A. (2014). Phonology. In J. R. Taylor & J. Littlemore (Eds.), *Bloomsbury companion to cognitive linguistics* (pp. 357–392). London: Bloomsbury.
- Mompean, J. A., & Lintunen, P. (2015). Phonetic notation in foreign language teaching and learning: Potential advantages and learners' views. *Research in Language*, 13(3), 292–314. <https://doi.org/10.1515/rela-2015-0026>
- Muller Levis, G., & Levis, J. (2016). Intonation bridging activities: Meaningful practice for final intonation. In J. Levis, H. Le, I. Lucic, E. Simpson, & S. Vo (Eds.), *Proceedings of the 7th pronunciation in second language learning and teaching conference* (pp. 229–235). Ames, IA: Iowa State University.
- Munro, M. J., & Derwing, T. M. (2008). Segmental acquisition in adult ESL learners: A longitudinal study of vowel production. *Language Learning*, 58(3), 479–503. <https://doi.org/10.1111/j.1467-9922.2008.00448.x>

- Murphy, J. (1991). Oral communication in TESOL: *Integrating speaking, listening and pronunciation*. *TESOL Quarterly*, 25(1), 51–75. <https://doi.org/10.2307/3587028>
- Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, 50(3), 417–528. <https://doi.org/10.1111/0023-8333.00136>
- Nyquist, J. D., & Wulff, D. H. (1996). *Working effectively with graduate assistants*. Thousand Oaks, CA: Sage.
- O'Connor, J. D. (1973). *Phonetics*. Harmondsworth, UK: Penguin.
- Okada, A. (2002). The usefulness of shadowing on prosody teaching of English. *Tsukuba International University Bulletin*, 8, 117–129. <https://doi.org/10.20843/00000136>
- Orth, J. L. (1982). University undergraduate evaluation reactions to the speech of foreign teaching assistants. *Dissertation Abstracts International*, 43, 3442A-4071A. (University Microfilms No. 83-09, 183).
- Oxford, R. L. (1990). *Language learning strategies: What every teacher should know*. Boston: Heinle & Heinle.
- Palmer, H. E. (1924). *A grammar of spoken English on a strictly phonetic basis*. Cambridge: W. Heffer and Sons.
- Pawlak, M. (2010). Designing and piloting a tool for the measurement of the use of pronunciation learning strategies. *Research in Language*, 8, 189–202. <https://doi.org/10.2478/v10015-010-0005-6>
- Pennington, M. C. (1989). Teaching pronunciation from the top down. *RELC Journal*, 20(1), 20–38. <https://doi.org/10.1177/003368828902000103>
- Pennington, M. C. (1996). *Phonology in English language teaching: An international approach*. London and New York: Longman.
- Pennington, M. C. (1998). The teachability of pronunciation in adulthood: A reconsideration. *International Review of Applied Linguistics*, 36(4), 323–341. <https://doi.org/10.1515/iral.1998.36.4.323>
- Pennington, M. C. (1999). Computer-aided pronunciation pedagogy: Promise, limitations, directions. *Computer Assisted Language Learning*, 12(5), 427–440. <https://doi.org/10.1076/call.12.5.427.5693>
- Pennington, M. C. (2015). Teacher identity in TESOL. In Y. L. Cheung, S. B. Said, & K. Park (Eds.), *Teacher identity and development in applied linguistics: Current trends and perspectives* (pp. 16–30). New York: Routledge.
- Pennington, M. C. (2018). Identity in language learning. In J. C. Richards & A. Burns (Eds.), *The Cambridge guide to learning English as a second language* (pp. 91–98). New York: Cambridge University Press.

- Pennington, M. C., & Hoekje, B. J. (2014). Framing English language teaching. *System*, 46(4), 163–175. <https://doi.org/10.1016/j.system.2014.08.005>
- Pennington, M. C., & Körmücü, S. (2018). Trialing a pedagogy designed to minimize L1 phonological transfer: Student and teacher responses to a curriculum for teaching Turkish first by ear then by eye. Unpublished ms.
- Pennington, M. C., & Richards, J. C. (2016). Teacher identity in language teaching: Integrating personal, contextual, and professional factors. *RELC Journal*, 47(1), 1–19. <https://doi.org/10.1177/0033688216631219>
- Pickering, L. (2001). The role of tone choice in improving ITA communication in the classroom. *TESOL Quarterly*, 35(2), 233–255. <https://doi.org/10.2307/3587647>
- Plonsky, L., & Mills, S. V. (2006). An exploratory study of differing perceptions of error correction between a teacher and students: Bridging the gap. *Applied Language Learning*, 16(1), 55–74. Retrieved August 8, 2017, from <http://www.dliflc.edu/wp-content/uploads/2014/04/all16one.pdf>
- Reed, M., & Michaud, C. (2015). Intonation in research and practice: The importance of metacognition. In M. Reed & J. M. Levis (Eds.), *The handbook of English pronunciation* (pp. 454–470). Hoboken, NJ: John Wiley & Sons.
- Richards, J. C. (2012). Competence and performance in language teaching. In A. Burns & J. C. Richards (Eds.), *The Cambridge guide to pedagogy and practice in second language teaching* (pp. 46–59). New York: Cambridge University Press. Abridged version of an article which appeared 2010 in *RELC Journal*, 41(2), 101–122.
- Rogerson, P., & Gilbert, J. B. (1990). *Speaking clearly*. Cambridge: Cambridge University Press.
- Rogerson-Revell, P. (2011). *English phonology and pronunciation teaching*. London: Bloomsbury.
- Ross, C., & Dunphy, J. (2007). *Strategies for teaching assistant and international teaching assistant development: Beyond micro teaching*. San Francisco: Jossey-Bass.
- Rubin, D. L. (1992). Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher Education*, 33(4), 511–531. <https://doi.org/10.1007/BF00973770>
- Rubin, D. L., & Smith, K. A. (1990). Effects of accent, ethnicity, and lecture topic on undergraduates' perceptions of nonnative English-speaking teaching assistants. *International Journal of Intercultural Relations*, 14(3), 337–353. [https://doi.org/10.1016/0147-1767\(90\)90019-S](https://doi.org/10.1016/0147-1767(90)90019-S)

- Saito, K. (2014). Experienced teachers' perspectives on priorities for improved intelligible pronunciation: The case of Japanese learners of English. *International Journal of Applied Linguistics*, 24(2), 250–277. <https://doi.org/10.1111/ijal.12026>
- Saito, K., & 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(2), 595–633. <https://doi.org/10.1017/S0272263112000356>
- Savignon, S. (1991). Communicative language teaching: State of the art. *TESOL Quarterly*, 25(2), 261–277. <https://doi.org/10.2307/3587463>
- Schmidt, R. W. (1983). Interaction, acculturation, and the acquisition of communicative competence: A case study of an adult. In N. Wolfson & E. Judd (Eds.), *Sociolinguistics and second language acquisition* (pp. 137–174). Rowley, MA: Newbury House.
- Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129–158. <https://doi.org/10.1093/applin/11.2.129>
- Schmidt, R. W. (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition*, 14(3), 357–385. <https://doi.org/10.1017/S027226310001189>
- Schmidt, R. W. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). New York: Cambridge University Press.
- Schumann, J. (1978). *The pidginization process: A model for second language acquisition*. Rowley, MA: Newbury House.
- Seidlhofer, B., & Dalton-Puffer, C. (1995). Appropriate units in pronunciation teaching: Some programmatic pointers. *International Journal of Applied Linguistics*, 5(1), 135–146. <https://doi.org/10.1111/j.1473-4192.1995.tb00076.x>
- Shiki, O. (2011). Exploring the relationship between shadowing instruction and L2 reading comprehension among Japanese university students. *Kwansei Gakuin University Humanities Review*, 15, 51–64.
- Sifakis, N. C., & Sougari, A. M. (2005). Pronunciation issues and EIL pedagogy in the periphery: A survey of Greek state school teachers' beliefs. *TESOL Quarterly*, 39(4), 467–488. <https://doi.org/10.2307/3588490>
- Sinclair, J. M. (Ed.). (2004). *How to use corpora in language teaching*. Amsterdam and Philadelphia: John Benjamins.
- Smith, J. A., Walters, A. J., & Burkhalter, C. M. (1992). *Communicate: Strategies for international teaching assistants*. Long Grove, IL: Waveland Press.
- Smith, K. (1993). A case study on the successful development of an international teaching assistant. *Innovative Higher Education*, 17(3), 149–163. <https://doi.org/10.1007/BF00915598>

- Smith, R. M., Byrd, P., Nelson, G. L., Barrett, R. P., & Constantinides, J. C. (1992). *Crossing pedagogical oceans: International teaching assistants in U.S. undergraduate education*. ASHE-ERIC Higher Education Report No. 8. Washington, DC: George Washington University, School of Education and Human Development. ERIC Document No. ED 358810.
- Spada, N. (1997). Form-focused instruction and second language acquisition: A review of classroom and laboratory research. *Language Teaching*, 30(1), 73–87. <https://doi.org/10.1017/S0261444800012799>
- Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning*, 60(2), 263–308. <https://doi.org/10.1111/j.1467-9922.2010.00562.x>
- Stevenson, I., & Jenkins, S. (1994). Journal writing in the training of international teaching assistants. *Journal of Second Language Writing*, 3(2), 97–120. [https://doi.org/10.1016/1060-3743\(94\)90010-8](https://doi.org/10.1016/1060-3743(94)90010-8)
- Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principles and practice in applied linguistics* (pp. 125–144). Oxford: Oxford University Press.
- Szyszka, M. (2015). Good English pronunciation users and their pronunciation learning strategies. *Research in Language*, 13(1), 93–106. <https://doi.org/10.1515/rela-2015-0017>
- Tamai, K. (2005). Strategic effect of shadowing on listening ability. In J. White (Ed.), *Proceedings of the 4th international conference on foreign language education and technology (FLEAT IV)* (pp. 620–625). Japan: Association for Language Education and Technology Kobe.
- Tarone, E., Bigelow, M., & Hansen, K. (2009). *Literacy and second language oracy*. Oxford and New York: Oxford University Press.
- Tergujeff, E. (2013). *English pronunciation teaching in Finland*. PhD dissertation published as Jyväskylä Studies in Humanities, 207. Jyväskylä, Finland: University of Jyväskylä. Retrieved September 23, 2017, from <http://tinyurl.com/nd624kz>. Google Scholar.
- Terrell, T. D. (1977). A natural approach to second language acquisition and learning. *Modern Language Journal*, 61(7), 325–336. <https://doi.org/10.1111/j.1540-4781.1977.tb05147.x>
- Terrell, T. D. (1982). The natural approach to language teaching: An update. *Modern Language Journal*, 66(2), 121–132. <https://doi.org/10.1111/j.1540-4781.1982.tb06970.x>
- Thomson, R. I. (2011). Computer assisted pronunciation training: Targeting second language vowels: Perception improves pronunciation. *CALICO Journal*, 28(3), 744–765. <https://doi.org/10.11139/cj.28.3.744-765>

- Thomson, R. I. (2013). ESL teachers' beliefs and practices in *pronunciation* teaching: Confidently right or confidently wrong? In J. Levis & K. LeVelle (Eds.), *Proceedings of the 4th Pronunciation in Second. Language Learning and Teaching Conference*, August 2012 (pp. 224–233). Ames, IA: Iowa State University.
- Underhill, A. (2010, October 6). *Pronunciation the poor relation?* British Council, Teaching English [Blog post]. Retrieved February 26, 2018, from <https://www.teachingenglish.org.uk/article/pronunciation-poor-relation>
- Underhill, A. (2012, August 28). Proprioception and pronunciation [Blog post]. Retrieved February 26, 2018, from <http://www.adrianunderhill.com/2012/08/28/proprioception-and-pronunciation/>
- Vannestål, M., & Lindquist, H. (2007). Learning English grammar with a corpus: Experimenting with concordancing in a university grammar course. *ReCALL*, 19(3), 329–350. <https://doi.org/10.1017/S0958344007000638>
- Wang, X., & Munro, M. J. (2004). Computer-based training for learning English vowel contrasts. *System*, 32(4), 539–552. <https://doi.org/10.1016/j.system.2004.09.011>
- Widdowson, H. G. (1978). *Teaching language as communication*. Oxford: Oxford University Press.
- Wilson, M. (2003). Discovery listening: Improving perceptual processing. *ELT Journal*, 57(4), 335–343. <https://doi.org/10.1093/elt/57.4.335>
- Wong, J. W. S. (2015). The impact of L2 proficiency on vowel training. In J. A. Mompean & J. Fouz-González (Eds.), *Investigating English pronunciation: Trends and directions* (pp. 219–239). Basingstoke, UK and New York, NY: Palgrave Macmillan.
- Wong, W. (2005). *Input enhancement: From theory and research to the classroom*. New York, NY: McGraw Hill.
- Wulff, S., & Ellis, N. C. (2018). Usage-based approaches to SLA. In D. Miller, F. Bayram, J. Rothman, & L. Serratrice (Eds.), *Bilingual cognition and language: The state of the science across its subfields* (pp. 37–56). Amsterdam & Philadelphia: John Benjamins.
- Yilmaz, E., & Soruč, A. (2015). The use of concordance for teaching vocabulary: A data-driven learning approach. *Social and Behavioral Sciences*, 191(2), 2626–2630. <https://doi.org/10.1016/j.sbspro.2015.04.400>. Retrieved February 26, 2018, from <http://www.sciencedirect.com/science/article/pii/S1877042815026609>
- Zielinski, B. (2015). The segmental/suprasegmental debate. In M. Reed & J. Levis (Eds.), *The handbook of English pronunciation* (pp. 397–412). Hoboken, NJ: John Wiley & Sons, Inc.



5

Using Technology for Pronunciation Teaching, Learning, and Assessment

Introduction

Interest in the use of technologies to facilitate language learning (**Computer-Assisted Language Learning**, or **CALL**) has grown rapidly in recent decades, and the market for CALL products, particularly in Asia, is enormous. The field of **Computer-Assisted Pronunciation Training**, or **CAPT**—including for both language teaching/learning and speech therapy—is also developing rapidly. Both are having major effects on education and language learning and teaching, and so research on their effects is important. In addition, both are becoming big business, which in some cases means that sales are running ahead of educational value.

There are obvious affordances of technology-based teaching and learning, such as the motivational effect created by the novelty or “wow” factor, the availability of multimodal resources incorporating audio, print, and video, and the enhanced mobility of being able to study anytime anywhere. CAPT can also provide endless opportunities for repetition and imitation, instantaneous responses, and exposure to a wide range of target language speech; it can also facilitate individualized, self-paced

learning. One of the main benefits of CAPT technology is the opportunity to provide automated feedback, and the use of speech technologies can be particularly beneficial for giving feedback on pronunciation. In addition, a number of the technologies used for teaching and learning pronunciation have obvious applications to research and testing.

In this chapter, we review the available technologies, how they are used, and a wide range of specific products and systems that are geared specifically for pronunciation teaching/learning or might be applied for pronunciation pedagogy, in addition to some computer-aided forms of assessment that incorporate pronunciation. We take a critical stance in our review that addresses issues of appropriate use and effectiveness, while also underscoring the exciting potentials of applications of speech technologies to pronunciation.

Speech Technologies for Pronunciation Teaching, Learning, and Testing

Speech technologies used in pronunciation teaching and learning, research, and assessment typically focus on speech analysis, speech recognition, and/or speech synthesis. **Speech analysis** enables the acoustic analysis of a speech signal, usually visualized as a waveform, speech contour, or spectrogram. Some speech analysis and display software is freely available, such as *Wavesurfer*¹ and *Praat*² (widely used as a research tool in phonology), and many commercial CAPT programs incorporate this feature. **Automatic Speech Recognition (ASR)**, also known as **speech-to-text (STT)** or voice recognition, converts the speech signal into words. ASR is an active area of linguistic and specifically phonological research as well as for the development of practical tools. Commercial ASR software, such as *Dragon Naturally Speaking* (or the older *Dragon Dictate*)³ is widely used for a variety of professional and personal communications (see Chap. 7). **Speech synthesis** refers to computer-generated speech. One type of speech synthesis is **text-to-speech (TTS)**, that is, automatic conversion of text into speech. Another type of speech synthesis employs computer programs designed to produce speech by simulating articulatory

movements or speech waveforms or by linking together digital representations of phonological segments and prosodies (sometimes called “synthesis-by-rule”). Speech recognition and speech synthesis systems can be combined to create dialogue-based systems, which are increasingly being used in some speech training applications. Again there are research and practical applications and both commercial and freely available programs. Progress in speech recognition and synthesis technologies in the last few years has made possible the development of computerized personal assistants, such as Apple’s *Siri*⁴ Amazon’s *Alexa*,⁵ and Microsoft’s *Cortana*,⁶ which can now process a wide variety of questions and give accurate answers in a natural-sounding voice.

Developments in educational technology have led to a wider range of delivery mechanisms, from more traditional companion websites and CD versions of books (e.g., *The Cambridge English Pronouncing Dictionary; Clear Speech*; Gilbert, 2012), to mobile applications or “apps” (e.g., *Pronunciation Power*,⁷ *Sounds*⁸) and resources on social media platforms such as *Twitter*, *Facebook*, and *WhatsApp*. Some commercial products, such as *Pronunciation Power*, are available in various formats, including DVD, interactive website, and mobile app. Such resources are appealing for pronunciation teaching and learning for many reasons, as enumerated by Pennington (1999) and as discussed more recently by Fouz-González (2015a, 2015b), especially given the constraints of the traditional language classroom. They have the potential to offer a private, stress-free environment where learners can access virtually unlimited input and practice at their own pace; and, with the integration of Automatic Speech Recognition (ASR), they can provide individualized, instantaneous feedback.

In addition to applications of the speech technologies described above in language teaching and learning, speech therapy (Chap. 7) makes extensive use of **electropalatography** (EPG), such as the *LinguaGraph*⁹ system and accompanying *LinguaView*¹⁰ software in which an artificial palate with embedded electrodes is fitted over a speaker’s upper teeth in order to identify tongue position and movement while articulating specific sounds or running speech (see Chap. 7). The pattern of tongue positions and movements are displayed on a screen as an aid in diagnosis, visual feedback, and research focused on problems in articulation of

consonants, which are often an area of developmental delay in children; of vowels in cleft palate; and other kinds of phonological disorders in children and adults (Gibbon & Lee, 2007).

Pedagogy vs. Technology

On closer examination, many commercially available CAPT materials are technology-driven rather than pedagogy-driven and lack the solid pedagogical grounding and practical guidelines required to sustain effective learning (Pennington, 1999). Levis (2007) claims that despite the fact that technology could fit a critical need in pronunciation teaching, “effective commercial CAPT applications are less innovative either in pedagogy or use of computer technology than one might expect” (p. 185). One of the difficulties is that there is no obvious fit between language learning pedagogies and the affordances of digital technologies. Computers are well suited for practice based on repetition, mimicry, or drilling, as advocated in audiolingual approaches, and for intensive one-on-one speech therapy, but it is less easy to apply these technologies within communicative methodologies. At present, limitations on **Artificial Intelligence (AI)** and speech recognition and synthesis mean that it is hard for a computer to really communicate or negotiate meaning with a human. Whether such limitations can be overcome eventually or are inherent limits to computer capabilities in respect of human intelligence and language remains to be seen.

A common criticism of CAPT, and CALL more generally, is that although technology has the potential to provide tailored, individualized feedback to learners, most products adopt a “one size fits all” approach (Derwing & Munro, 2015). Given that technology often takes precedence over pedagogy, with the result that many resources often have a limited curriculum focus (typically, on phonemes), a limited range of activity types (e.g., minimal pair discrimination), and limited feedback (e.g., “correct” or “incorrect”). The novelty value of “drag-and-drop” or “odd-one-out” activities soon wears off unless supported by pedagogically sound feedback or support. As Neri, Cucchiari, Strik, and Boves (2002) point out, “many authors describe commercially available programs as

fancy-looking systems that may at first impress student and teacher alike, but eventually fail to meet sound pedagogical requirements” (p. 442). In other words, it is essential that a concern for technological affordances or innovation not be at the expense of sound pedagogic practice (Rogerson-Revell, 2011; Pennington, 1999). Some of Pennington’s (1999) principles for the design of CAPT materials are that CAPT systems should:

1. Establish baseline, reference accents for instruction;
2. Set measurable goals and performance targets;
3. Be designed to build skills from easier to more challenging exercises;
4. Link pronunciation to other aspects of communication; and
5. Raise users’ awareness of how their L1 phonological systems differ from the system of the target language. (p. 434)

Some CAPT resources have been developed through academic/technological collaborations and are well grounded in both pedagogic and language learning theory and practice, so ensuring greater relevance and validity of phonological content than some of the more commercially driven counterparts. A good example is Protea Textware’s *Connected Speech*,¹¹ which focuses on connected speech and provides multiple speakers of Australian, American, and British English, using ASR to provide feedback on the learner’s production. Cauldwell’s *Cool Speech* app¹² is similarly underpinned by extensive academic research and includes many features of fluent natural speech in British or American English. Similarly, Thomson’s *English Accent Coach*¹³ is based on research referred to earlier in high variability perceptual/phonetic training, or HVPT (Chap. 3). There is also a vast amount of pronunciation material available online, much of which is free to use, although as with any unedited materials, some caution is needed in selecting these websites. Again, some of the most professional and information-rich resources available have been developed by pronunciation specialists, such as *Sounds of Speech*¹⁴ (University of Iowa) or the *Phonology and Phonetics Review*¹⁵ (Rogerson-Revell, 2010) and *Web Tutorial*¹⁶ (University College London). There are also some useful sites created and maintained by enthusiastic CALL and CAPT specialists (e.g., Brett¹⁷; Powers¹⁸). Brett’s multimedia, interactive resources are particularly engaging both for teachers and students. Various studies have demonstrated the flexibility and value of

computer-based training for working on aspects of pronunciation, for instance, vowels and consonants (Neri, Cucchiarini, & Strik, 2006; Wang & Munro, 2004), rhythm and stress-timing (Coniam, 2002); intonation (Hardison, 2004; Kaltenboeck, 2002; Levis & Pickering, 2004); and speech rate and fluency (Hincks, 2005).

Feedback

A key challenge in technology-based language learning is providing adequate interactivity and intelligent, customized feedback, and it is these areas that are still particularly problematic for CAPT. It is relatively easy for a computer to provide feedback on a learner's perception of pronunciation, but providing accurate feedback on pronunciation in production presents challenges, especially outside a one-on-one tutorial situation in which a phonetics expert, such as a speech therapist, is available to provide customized coaching. Many programs and apps have been developed which provide simple feedback on whether a learner's perception of individual sounds (e.g., *Sounds* app), word stress (e.g., *Pronunciation Power*), or even a speech phrase (e.g., *Duolingo* app¹⁹) is correct or not. However, as Levis (2007) states, "Technologically, CAPT systems often suffer from difficulties in giving learners adequate, accurate feedback and an inability to provide accurate and automatic diagnosis of pronunciation errors" (p. 185).

Visual Feedback

Regarding feedback on pronunciation in production, one of the easiest ways is by some form of visual representation. For instance, spectrographic and other forms of acoustic analysis have been used to give visual feedback on intonation patterns since the 1960s. Some of the available systems, such as *VisiPitch IV*²⁰ (KayPentax) and *WinPitchLTL*²¹ (Germain-Rutherford & Martin, 2002) were originally designed for analysis by phoneticians and speech scientists rather than for pronunciation training; and while they can provide detailed phonetic information, it requires

some expert knowledge to manipulate and interpret the results and to help students or other clients use those results to improve their performance. VisiPitch IV allows learners to see their prosodic contour superimposed above or below that of another speaker, such as a tutor or a recording, which can serve as a model. The learner can then make further attempts to match the shape of the model contour. Creative games provide motivational practice opportunities. WinPitchLTL also allows the teacher or therapist to modify individual learners' speech signal so that they can hear and see the correct prosodic contours with their own voice, which has been shown to help the learner to perceive significant deviations (Nagano & Ozawa, 1990). However, the effectiveness of such auditory and visual input may depend on the availability of a teacher or therapist with specialized phonetic knowledge to interpret that input for the learner.

Some commercial CAPT systems incorporate visual feedback to be used for independent study, such as *Pronunciation Power*, which includes acoustic analysis for learners to compare their own recorded pronunciation of a sound with a model instructor's voice in the program. In this screenshot (Fig. 5.1), the sound being practiced is the long vowel /i:/ in General British pronunciation (represented as *iy*, a common way of representing this vowel phoneme for American English).

Some systems use visual displays of pitch contours rather than acoustic information. For instance, the *BetterAccentTutor*²² software was developed to teach American English prosody to non-native speakers. It provides visual feedback on intonation, stress, and rhythm, matching the learner's input with a native speech model. It is claimed that visual displays of pitch contours help learners improve both perceptual and productive accuracy of prosody (de Bot & Mailfert, 1982; Hardison, 2004). Moreover, Hardison's (2004) study of learners of French found that the audiovisual feedback improved not only their prosody but also their segmental accuracy. Creative display of pitch contours or other prosodic information such as rhythm and intensity can also be effective. For example, Hincks and Edlund (2009) developed software that employed a novel way of giving feedback to Chinese learners of English using flashing lights rather than acoustic representations to portray prosodic

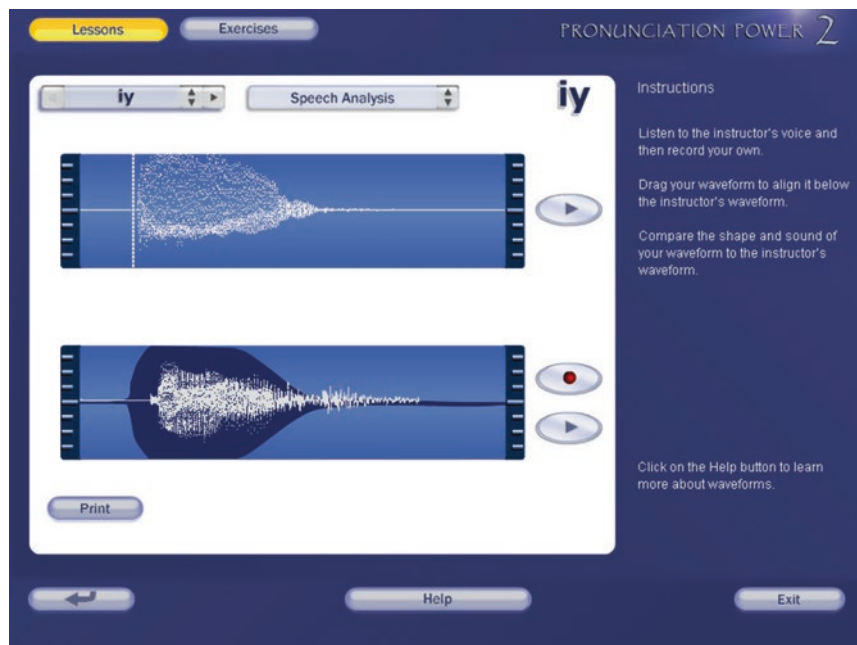


Fig. 5.1 Screenshot of acoustic feedback in *Pronunciation Power*

structure and found the feedback helped the students to speak with more liveliness.

While pairing visual feedback with auditory feedback has benefits for pronunciation training, some issues remain with these types of displays. First is the concern that it is not evident that learners can use such visual displays independently of a teacher with some expert knowledge. Secondly, it has been suggested that the improvements in pronunciation seen after using such displays may simply be due to learners spending more time practicing (de Bot, 1983). Another issue is that the technique of comparing the learner's speech signal with a model native speaker signal suggests that the aim of the training is for the learner to achieve as close a match as possible to the model speech signal. However, as Neri et al. (2002) point out, "In fact, this is not necessary at all: two utterances with the same content may both be very well pronounced and still have waveforms or spectrograms that are very different from each other"

(p. 453). Finally, visual displays cannot give specific feedback on the precise error or cause of the error, so one concern is that the learner may make random attempts at correction which ultimately may result in fossilized errors or demotivation (Eskenazi, 2009).

In sum, there appears to be considerable value in giving visual feedback to support pronunciation learning, although there may be limitations to the benefit of spectrograms and waveforms, at least when used as the only form of feedback (Ehsani & Knodt, 1998; Neri et al., 2002). It could be argued that pitch contours are more iconic and therefore more easily relatable to the rises and falls in a speaker's voice. They appear to be useful for intonation training (Hirata, 2004; Levis & Pickering, 2004), if some guidance is given in how to interpret them (Chun, 2013; Hardison, 2004). However, there is still some debate about which intonation features are most important for intelligibility (Chun, 2013), and so which features should be emphasized in giving feedback to learners.

An interesting alternative approach is resynthesizing the learner's speech input so that it sounds more nativelike. In resynthesis, the pitch and duration parameters of a native speaker are applied to a language learner's speech sample, such as in the *WinPitchLTL* system. A study by Felps, Bortfeld, and Gutierrez-Osuna (2009) suggests that resynthesis reduces the perception of a foreign accent while maintaining the individual voice quality properties of the speaker. A study by De Meo, Vitale, Pettorino, Cutugno, and Origlia (2013) demonstrated that for training Chinese speakers to use Italian prosodic patterns, self-imitation using resynthesis to technologically "transplant" improved intonation and other prosodic features onto a speaker's own voice was more effective than imitating a standard model. As the authors speculate about future efforts to develop speech resynthesizing techniques further for language learning:

By means of the prosodic-intonational transplantation procedure, the suprasegmental features of the native speaker (pitch, intensity, articulation and speech rate, frequency and duration of pauses) would be cloned and transferred in real time to the L2 learner's voice, without altering the perception of the L2 speaker's identity. The learner's voice thus becomes the "native" model to imitate. However, speakers with different L1s and

different levels of L2 competence, and a greater number of speech acts have to be considered in order to get sufficient data to support the development of a technological tool that makes teaching and/or autonomous learning of the L2 suprasegmental features easier. (De Meo et al., 2013, p. 98)

A text-to-speech app called *Voice Changer*²³ which allows users to convert text to speech with variable prosodic and voice quality features (see also Chap. 7) may have value for motivating learners' interest and focusing their attention on these aspects of speech. If learners like to use the app, it could help develop awareness of different styles and identity-linked aspects of perceptual competence in pronunciation. This kind of program can also provide a basis for pronunciation practice to imitate the different speech models provided, thus having the potential to influence production as well. In addition, there is perhaps future pedagogic potential in marrying this app in some way to the resynthesis technology investigated by Felps et al. (2009) and De Meo et al. (2013).

ASR Feedback

The area of speech technology which offers most potential towards the goal of individualized immediate feedback is Automatic Speech Recognition (ASR). ASR technology works by comparing speech input from one speaker with a model of native speaker speech, generated from a database containing recordings of hundreds of native speakers of the language. Earlier versions of ASR were speaker-sensitive and had to be trained to recognize individual speaker input, but more recent systems are less speaker-sensitive and can handle input without training on each user's individual voice. Indeed, progress in ASR technology in recent years has made it possible, for example, to use voice applications such as Apple's Siri in smartphones and Microsoft's Cortana in computers with little or no initial training necessary to recognize the user's voice, as long as it is an accent on which the device was trained.

However, further technological advances are required in AI, in terms of natural language processing and computational power, before computers can converse with human speakers in unconstrained, naturalistic

contexts. The difficulties are compounded when dealing with accented, non-native speech since such variations in pronunciation cannot easily be represented in ASR databases. Hence, while ASR programs are becoming increasingly effective for native speaker voice recognition, the accuracy level for non-native speech is much lower. Levis (2007) describes, for example, the ASR word processing application, *Dragon Naturally Speaking*²⁴ as “95% or more accurate for native-speaking English users” (p. 192), as claimed by the developer and some users (e.g., see Pogue, 2004), though some native speaker users report significant problems when speaking at normal speed or using technical terms with *Dragon* (Altman, 2013) or other voice recognition software (Hill, 2013). However, as Levis (2007), goes on to say, “accuracy for these programs, when used by advanced proficiency but accented nonnative speakers of English, drops to near 70%” (p. 192), as confirmed in studies by Coniam (1999) and Derwing, Munro, and Carbonaro (2000). Many of the difficulties of using voice recognition software for language learning were known or anticipated some twenty years ago by Ehsani and Knodt (1998).

Despite such limitations, ASR has been used in CAPT software since the 1990s, with the aim of providing individualized feedback based on an analysis of a learner’s voice input to the computer as a digital recording. In a typical ASR-based CAPT program, a learner will receive a prompt and will then select a response from a restricted set of possible responses (such as multiple choice) or that can only be answered in very limited ways. Such constrained responses mean that even accented speech is likely to be analyzable. The learner’s recorded response is then acoustically aligned with the speech model in the database and shows which sounds deviate most from that model. A numerical score can be given to show how much the speaker’s input matches or deviates from the stored model, but current technology cannot determine in what way the signal has deviated from the model.

The most popular CAPT systems are those that incorporate ASR, such as *Pronunciation Power*²⁵ (EnglishLearning) and *Tell Me More*²⁶ (Rosetta Stone), which are able to provide users with immediate and personalized (though not always correct) feedback. Both of these programs make extensive use of multimedia to provide interactive and engaging

pronunciation learning environments, covering both segmental and suprasegmental features. Rosetta Stone's *Tell Me More* language learning software uses ASR to analyze a learner's oral response to a multiple choice question on screen and, depending on the choice made, the dialogue will develop in a particular way. If the learner is not understood, he or she will be prompted to repeat the response. In this sense, the learner is involved in fairly realistic interactions and is made aware of the need for accurate pronunciation in order to progress. *Pronunciation Power* offers a "Speech Test" using ASR to assess students' recordings and to analyze which lessons they should study to improve their pronunciation.

According to the *Pronunciation Power* website,²⁷ "Once all the 52 test sentences have been recorded the program automatically analyzes the recordings and provides feedback on which sounds are spoken correctly, which need improvement, and greatly need improvement. The students are then directed to which lessons and exercises in the *Pronunciation Power* products they should focus on." Having tried the demo test, Pamela was informed that most of her sounds were "correct," but that two "needed improvement" and one "greatly needed improvement." Presumably, this was the result of a native GB speaker taking a test modelled on GA, showing the limitations of pronunciation programs given the diversity of accents, especially for such a widely spoken language as English.

Pronunciation Power also has a "Speech Analysis" area where users are able to record their own production of individual sounds, words, and/or complete sentences and then compare their waveform with that of the instructor. The waveforms provide information concerning the loudness (amplitude) and pitch (frequency) of sounds, as well as duration (length). The *Tell Me More* software (Auralog²⁸) includes a similar "listen-record-playback-compare" feature with its voice recognition feature, S.E.T.S. (Spoken Error Tracking System). This also provides an automated scoring system, rather than having to rely on a teacher to interpret the visual display. In this software, the computer program compares a stored model pronunciation with a learner's utterance and produces a global pronunciation score as well as color-coding words that are incorrectly pronounced within a sentence (see Neri et al., 2002).

However, the *Tell Me More* system appears to have difficulties handling a range of accents and differences in speech rate, even for native speakers, as one user review of the Spanish language version stated:

... I found the voice recognition software, S.E.T.S (Spoken Error Tracking System), to be incredibly troublesome. I am a native speaker of Spanish and found that even my voice was not recognized as it should have been. When a program does not allow for accents or even a slightly different speed of speech, then it isn't working as it should. You can say a word the exact same way as the CD speaker, and it will still accuse you of speaking too slowly or incomprehensibly. If this is the only connection that you have to the language you are trying to learn, it can be very disheartening. (user review 14, Feb 11, 2017, *Tell Me More Reviews*²⁹)

ASR errors can not only frustrate and mislead the learner but also, and more importantly, undermine their trust in the CAPT tool (Levis, 2007; Wachowicz & Scott, 1999). Unfortunately, erroneous feedback is not uncommon in CAPT software, particularly when incorporating ASR technology, as seen in evaluations of other programs, such as *TriplePlayPlus* (Mackey & Choi, 1998) and *Learn German Now!* (Wildner, 2002). Neri et al. (2002) explain that CAPT systems can be tuned to lean in favor of falsely accepting incorrect responses rather than falsely rejecting correct ones. They also advise that the amount of corrective feedback should be limited to avoid demotivating learners and to prioritize feedback on sounds with a high functional load. Neri et al. (2002) also suggest that errors should be prioritized based on how frequent and persistent the errors are, how perceptually relevant they are in terms of intelligibility, and how reliably they can be detected by ASR.

There is obvious potential in technology of providing immediate feedback on a user's oral production, for instance, through ASR and automatic scoring. If carefully designed, ASR feedback has been found to give positive results with phonemic training, for example, in teaching learners how to produce the /x/ sound in Dutch (Cucchiari, Neri, & Strik, 2009). However, precise feedback is difficult for suprasegmentals, as ASR is poor at handling information about a speaker's prosody, which means that ASR can give better feedback at the segmental level than at the

suprasegmental level (Levis, 2007). Levis (2007) summarizes the key components of effective computer-based pronunciation feedback based on the recommendations of a number of researchers, maintaining that it needs to:

- be consistent with human feedback;
- be immediate;
- be pertinent and correct;
- be given in a form that students can make use of;
- include information about when goals have been reached; and possibly
- suggest ways to address errors. (pp. 192–193)

Despite advances in ASR, the technology is still limited for pronunciation in terms of the accuracy of feedback and in helping learners locate and remedy specific errors on their own.

CAPT for Young Learners

The growing emphasis on multilingualism in many national educational policies is leading to the introduction of foreign language learning from an earlier age, for instance, in Italy, where learning a foreign language has been made compulsory from the first years of primary education. As a result, there is an increasing demand for speaking and pronunciation resources for young learners. Some of the affordances of computer technologies, particularly the opportunities for task-based, multimodal, interactive learning, can be beneficial for pronunciation training for this age group (Mich, Neri, & Giuliani, 2006).

The use of computers for educational purposes with children has increased rapidly in the last decade, and this includes language learning and pronunciation skills. The British Council's *Learn English Kids*³⁰ website, for example, has a specific "Pronunciation Activities" section. CAPT programs have also been developed which target either young adults or children, such as *English for Kids* (Krajka, 2001). Some of these involve task-based activities, including speech-based roleplays and games, such as

Auralog's³¹ *Tell Me More/Talk to Me Kids* series. There are also many resources aimed primarily at L1 pronunciation acquisition or remedial learning but which can be used for L2 learning. Typically, these focus on phonics and spelling, such as Innovative Investment's *Phonics Studio* app available in iTunes³² or the Commonsense Media *C is for Cow* app,³³ which teach children to recognize both the sound and look of each letter in the English alphabet. A similar app, *Articulation Station*,³⁴ includes voice recording and auditory feedback functions along with the ability to store and track a learner's progress.

There is growing interest in CAPT systems including ASR technology, such as *Ya-Ya Language Box* (Chou, 2005) and *Tell Me More Kids*, because of the opportunities for incorporating speaker input, especially through task-based, interactive activities. Some of these systems also provide automated feedback in some form, at the level of the phoneme, word or sentence. However, the accuracy of such feedback is still limited, partially because of the technical difficulties of recognizing and evaluating children's non-native speech, compared to adult native speech (Gerosa & Giuliani, 2004; Hacker et al., 2005). The risk of providing inaccurate or inappropriate feedback is a major issue with all ASR-based CAPT programs but is a particular concern for young learners. It is obviously beneficial to provide corrective feedback at an early stage of learning and thereby avoid fossilization of errors. Providing incorrect feedback to children can however have long-term consequences for pronunciation as well as motivation and engagement with language learning.

Research evaluating CAPT for young learners is relatively sparse and it is not easy to find studies which provide evidence of the effectiveness of such programs for the development of pronunciation skills in language learning. Neri, Mich, Gerosa, and Giuliani (2008) suggest that:

...no empirical data have been collected, to our knowledge, on the actual pedagogical effectiveness of these systems for children. Research seems to be driven more by technological development rather than by the pedagogical needs of learners. As a result, systems with sophisticated features are built and sold, but we do not know whether the features and functionalities that they include will actually help learners to achieve better pronunciation skills. (p. 395)

Some studies have shown that children do enjoy using computers for pronunciation training (Chou, 2005; Mich et al., 2006). Moreover, Neri et al.'s (2008) study of a CAPT program for young Italian learners of English indicated that "training with a computer-assisted pronunciation training system with a simple automatic speech recognition component can lead to short term improvements in pronunciation that are comparable to those achieved by means of more traditional, teacher-led pronunciation training" (p. 393). Such studies demonstrate the potential benefits of technology for pronunciation instruction when well designed and used appropriately.

Some Recent Developments

Rapid advances in digital technologies and online resources, including in ease of use and access, has led to the proliferation of digital media for communication, entertainment, and education. Within an increasingly globalized, multilingual world, knowing how to use these technologies and resources to learn languages is important for learners as well as teachers.

Mobile Apps

The use of mobile devices, such as smartphones and tablets, is a growing trend in language learning generally, including CAPT. Mobile apps not only offer the affordances of convenience and ease of access, but also the personalization and localization of mobile-assisted language learning, or **MALL** (Díaz-Vera, 2012). As technological advances lead to improved processing power, screen compactness and visibility, and network connections, so the communicative and media production capabilities of mobile devices expand, offering an optimal environment for multimodal, communicative language learning, including pronunciation training (Fouz-González, 2012). A large number of language learning apps are now available, including pronunciation-specific apps such as Oxford's *English File Pronunciation*,³⁵ and research is being carried out on their use. For instance, Fouz-González's (2015a) research with L2 English students from

different fields of study who used the Oxford *English File Pronunciation* app on their smartphones demonstrated its potential to improve both perception and production of vowels. As with most online resources, there is a great range in quality and educational value among the available apps, although some of the best ones have been developed from existing pedagogically sound resources, such as Macmillan's *Sounds* app,³⁶ Cambridge's *Clear Speech* app,³⁷ and the *Cool Speech* iPad app.³⁸

Given the potential of ASR to provide customized feedback on pronunciation input, it is not surprising that there has been considerable interest in this area by mobile app developers. The increasing emphasis on AI in smartphones is enabling the development of advanced apps which act as so-called "intelligent" language tutors. Apps such as *Duolingo*³⁹ or the *ELSA Speak*⁴⁰ and *Supiki*⁴¹ apps use voice recognition and analysis software to give individualized feedback on a user's speech recording and have controlled interactions in the way of structured conversations with the user.

Chatbots

The goal of having a natural, spontaneous conversation with a computer is a distant one; however, rapid advances in speech technologies mean that it is increasingly possible to have useful exchanges with machines. A "chatterbot," or **chatbot** is a computer program which uses AI to simulate a conversation with a human interlocutor either by audio or text. The first chatbot was a computer program developed in the 1960s, *ELIZA* (Weizenbaum, 1966) which was designed to simulate a psychiatric interview between doctor and patient taking place as an interaction between the computer and someone typing in English words. The software appeared to understand and be able to participate in authentic interactions by giving programmed responses to keywords and phrases. Since then, advances in machine learning and computing power has led to the creation of many more chatbots, such as *Jabberwacky*⁴² which can be used freely on the internet or downloaded onto a mobile device. Fryer and Carpenter (2006) claim that *Jabberwacky* can be used successfully for language learning purposes, particularly as it incorporates text and synthesized speech, allowing

students to practice both listening and reading skills. They describe the potential educational value of bots, concluding that “these positive communicative experiences with chatbots could create new or renewed interest in language learning and improve students’ motivation” (p. 10).

Most chatbots were designed to interact with and entertain native speakers, so initially they were most useful to advanced level language learners. However, recently, chatbots have been developed specifically for language learners, including beginners. For example, in the *Duolingo* app, bots are currently available for learning French, Spanish and German, but only for iPhone users. The developers plan to extend the range of languages offered and also to enable speech-based, as well as text-based, chats.

The *Duolingo* Spanish bot is text-only at present (it talks to the user, but the user has to type responses), but the plan is to introduce a fully spoken version in the future. An app named *Mondly*⁴³ (ATi Studios) has recently been launched which combines voice recognition with recorded speech and adaptive visual responses to simulate intelligent conversational exchanges. The developers of *Mondly* claim that the “Conversational Chatbot” provides “fun and adaptive lessons that encourage users to practice the language they are learning in everyday scenarios, such as ordering in a restaurant. The app recognizes millions of inputs and creates an adaptive visual response when it recognizes a word or phrase that the user has said, providing a reinforcing feedback that helps build confidence”.⁴⁴ While pronunciation practice at present is limited in such apps, if speech synthesis can be included, chatbots open up vast potential for conversation and pronunciation practice with an online intelligent tutor.

Embodied Conversational Agents or “Talking Heads”

Chatbots represent a key step in the development of CALL and CAPT programs to be able to use spoken dialogue systems for language speaking practice. A spoken dialogue system combines speech recognition, natural language understanding, and speech synthesis to enable a person to communicate with a computer through voice and complete a task. Developers are working on **Embodied Conversational Agents** (ECAs) that can act as both language tutors and conversational partners.

Research over more than a half century has shown that the face presents visual information during speech that supports effective communication and improves intelligibility (Benoît, Mohammadi, & Kandel, 1994; Hardison, 2007; Jesse, Vrignaud, Cohen, & Massaro, 2000; Sumbly & Pollack, 1954). ECAs, or virtual 3D **talking heads** are increasingly being used in research and applications in speech science. An example is *Baldi*⁴⁵ a talking head developed at the Perceptual Science Laboratory at the University of California at Los Angeles (see Fig. 5.2).

Ouni et al. (2005) explain the images in Fig. 5.2 as “Three Different Views: In the middle, the standard Baldi; to the left, semi-transparent Baldi (which allows to see the inner articulation: tongue, palate and teeth); to the right, the wire frame” (p. 116).

Baldi has been found to be particularly effective for communication with the hard of hearing (Massaro & Light, 2004), who are able to view the interior and exterior physical gestures involved in speech, and has been used as a vocabulary tutor with autistic children (Bosseler & Massaro, 2003). *Baldi* can also function effectively as a spoken language tutor, a reading tutor, or personal agent in human–machine interaction (Ouni et al., 2005, p. 116). Ouni et al. (2005) describe their research to extend *Baldi*'s capability beyond English to become a multilingual talking head able to speak a variety of additional languages, including development of an Arabic talking head, *Badr*. The development of a multilingual talking head, if successful, will facilitate

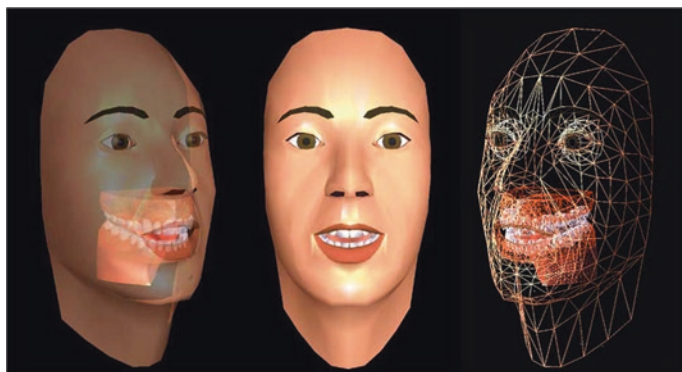


Fig. 5.2 Talking head *Baldi* (Ouni, Cohen, & Massaro, 2005)

research into phonological similarities and differences between languages and will have many potential applications, including for second language pronunciation instruction.

Alsabaan and Ramsay (2014) developed a talking head for use as a diagnostic tool to help L2 learners of Arabic improve their pronunciation. The talking head provided feedback showing the articulations which the learners made and the sounds which they should have made, as well as an explanatory text of how they could pronounce the target sample correctly. The study involved 40 students of Arabic, and preliminary results from this pilot group indicated that the students' pronunciation improved over the course of the study.

ECAs have been used in other CAPT contexts. Wik and Hjalmarsson (2009) describe the development of two animated talking heads, *Ville* and *DEAL*, which have both been designed for language learning applications but with different roles and functionality. They describe *Ville*⁴⁶ as “a virtual teacher’ whose role is to guide, encourage, and give corrections on a student’s pronunciation and language use” (Wik & Hjalmarsson, 2009, p. 1025). Users can try perception and production exercises (see Fig. 5.3), and the software uses a contrastive analysis approach to correct typical errors of L2 Swedish learners.



Fig. 5.3 *Ville* screenshot—Pronunciation exercise

DEAL,⁴⁷ on the other hand, is seen as a “role-play dialogue system for conversation training” (Wik & Hjalmarsson, 2009, p. 1025). Wik and Hjalmarsson explain that the systems differ in the types of feedback they offer: whereas *Ville* focuses on segmental accuracy, *DEAL* focuses on intelligibility. *DEAL* does not comment on a learner’s performance, but “acts as a conversational partner” (Wik & Hjalmarsson, 2009, p. 1025) giving feedback in the form of backchannels or clarification questions where necessary (see Fig. 5.4).

Wik and Hjalmarsson (2009) compare *DEAL* and *Ville* as follows:

DEAL serves as an important complement to *Ville*; whereas *Ville* provides exercises on isolated speech segments, i.e. phone, syllable, word, and sentence level, *DEAL* adds the possibility of practicing these segments in the context of a conversation. *Ville* has the role of a teacher who gives you feedback and help when you encounter problems. *DEAL* on the other hand has the role of a native speaker, for example, a person with a service occupation, whom you need to communicate with using your new language. (p. 1033)

A similar approach is taken in other resources. For instance, in *TraciTalk*⁴⁸ (Harashima, 1999; Wachowicz & Scott, 1999), a system which was conceived as a more generic CALL environment rather than a CAPT system, the learner interacts with an ECA whose task is to help the learner to solve a mystery using the target language.

Gaming and Simulations

The gaming element of programs such as *TraciTalk* and *DEAL* adds a further level of motivation and engagement to language learning, allowing users to practice oral skills in a fun and challenging context. The affordances of online games and simulations for language learning are increasingly being recognized both for CAPT and for CALL more generally (Golonka, Bowles, Frank, Richardson, & Freynik, 2014). This parallels the rapid growth in multiplayer online gaming and mobile games for smartphones, particularly (but not exclusively) for young people. As Godwin-Jones (2014a) observes, “If language learning can be tied

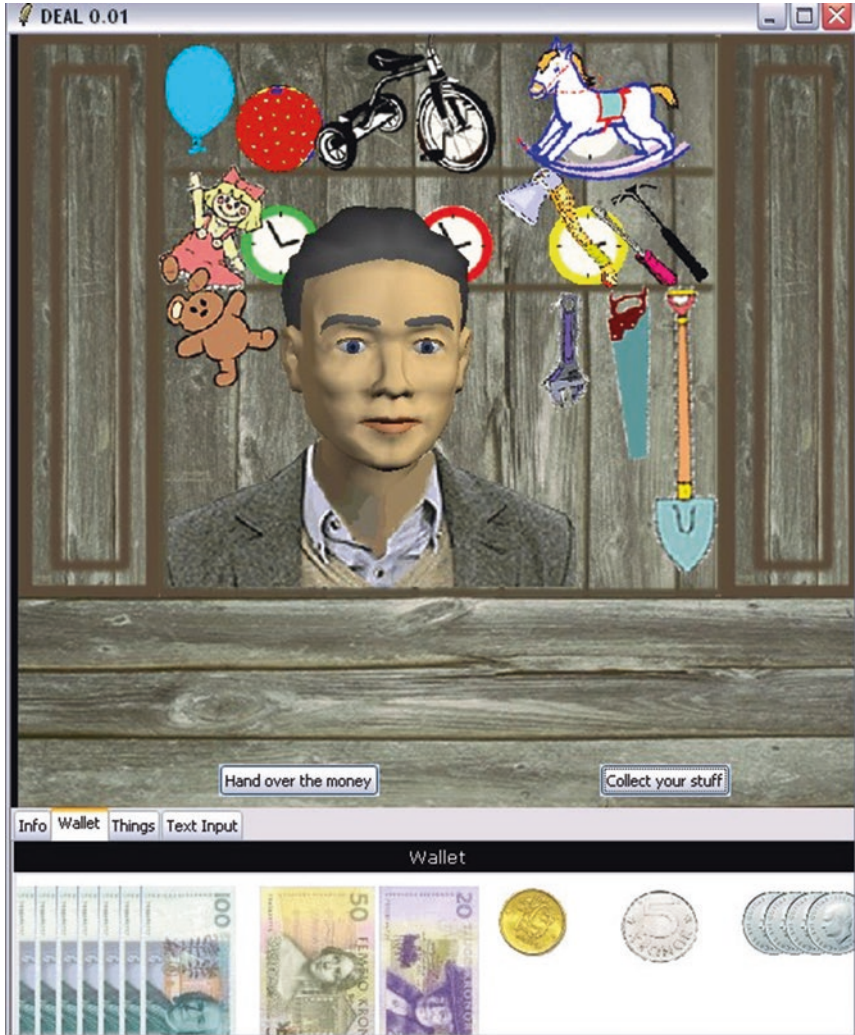


Fig. 5.4 DEAL screenshot

to popular forms of gaming in a way that does not inhibit its enjoyment, that's a winning situation both for students and educators" (p. 9). The elements of competition, problem solving and reward for task completion can be particularly motivating, while the use of roleplays or avatars

can depersonalize communication and help reduce stress. Games can also provide an immersive environment involving extensive use of the target language. To make progress in a game, players typically have to interact verbally with game objects or other players, so they use language in real and meaningful ways to accomplish a task. The emphasis in such game contexts is on language which is intelligible and socially appropriate rather than necessarily 100% accurate. Gamers often interact with other players from a wide range of cultural and linguistic backgrounds and need to be able to understand and respond to a variety of language input quickly and effectively. All of these aspects of games can be of benefit for teaching and learning pronunciation.

The most popular massively multiplayer online game (**MMOG**) on the market today is *World of Warcraft (WoW)*, which has over 12 million users and is available in multiple languages. As with other MMOGs, players advance through the game scenario and gain game-playing skills by completing quests, collecting or making items, and buying and selling goods or services. Although commercial off-the-shelf (**COTS**) games such as *World of Warcraft* have not been designed for language learning, a number of studies have examined *WoW*'s language learning potential (Nardi, Ly, & Harris, 2007; Rama, Black, Van Es, & Warschauer, 2012; Thorne, Fisher, & Lu, 2012), which—with some creative pedagogy on the part of the teacher—may include application to the development of speaking and pronunciation skills.

Some attempts have been made to modify COTS games or develop “serious” games, such as games for education or language learning. An example is *Croquelandia*⁴⁹ (Sykes, 2013), created specifically with the goal of enhancing learners’ ability to perform requests and apologies in Spanish. Other language learning games have been developed for military training purposes, such as the U.S. Army’s Military Language Trainer (MILT; Holland, Kaplan, & Sabol, 1999). MILT includes a microworld in which a speech recognition system allows a learner to solve a problem by using the target language (Arabic) to manipulate an animated agent searching a series of rooms. Results of a pilot study concluded that “one hour’s use of the MILT Arabic tutor significantly improved students’ Arabic proficiency on four dimensions of language usage—vocabulary, grammar, pronunciation, and overall fluency and that students’ attitudes

toward the tutor were favourable” (Holland et al., 1999, p. 9). A similar, more recent, development is the Tactical Language and Culture Training System (TLCTS), designed for use by U.S. military personnel (Lewis Johnson, 2010). TLCTS uses a variety of technologies, including natural language processing, speech recognition, and artificial intelligence agents, to provide simulated encounters with native speakers in the target culture. Multiple TLCTS training systems have been developed to date including *Tactical Iraqi*⁵⁰ (Fig. 5.5), which teaches Iraqi Arabic language and culture by means of simulated encounters with people in an Iraqi context (Lewis Johnson, 2010).

TLCTS gives priority to conversational practice but includes feedback on pronunciation. A new version (5.0) of *Tactical Iraqi* includes a wider



Fig. 5.5 Screenshot from *Tactical Iraqi*

range of missions and new lessons focusing on common problems in Arabic pronunciation, to help learners get a better grounding in the sound system of Arabic. An internal study of *Tactical Iraqi* showed overall substantial gains in language proficiency by those who used this software (Surface & Dierdorff, 2007).

Robot-Assisted Language Learning

Not long ago, Han (2012) made the following prediction: “Just as many people now have a personal computer, in the near future, personal robots (PR) may become the next paradigm-shifting tool for everyday life” (p. 8). Robots have been used since the 1960s in manufacturing industries, such as factory automation in the automotive industry. Today, robots are being used in a much wider range of applications and functions, such as in vacuum cleaners, lawn mowers, autonomous vehicles, and as tour guides, care home assistants, and hotel receptionists. Their use in education is still relatively novel; but research and development of robot-assisted language learning, or **RALL**, began in the mid-2000s, mainly in such “expanding circle” (Kachru, 1985) countries as Japan, Korea, and Taiwan, where English is taught as a foreign language. In such countries, it is common to employ L1 English-speaking teaching assistants to help young learners with their pronunciation learning, often in after-school programs. However, if it is not possible to employ native speakers for face-to-face instruction, classes in these countries may be conducted with computer, video, or mobile-based applications. The use of robots is seen as an alternative pedagogical approach, in which the robot can take the role of the native speaker and interact with the learners.

According to Han (2012):

Among the various instructional models in language learning, we should consider RALL, employing currently emerging robot technology. This anthropomorphized version of existing mobile devices is autonomous, with features such as image recognition through camera, voice recognition through microphone, and interaction based on various sensors. (p. 5)



Fig. 5.6 RALL with *IROBIQ* (left) and RALL with *ROBOSEM* (right) (Park, Han, Kang, & Shin, 2011)

Han illustrates the use of robots in English classes in South Korea as shown in Fig. 5.6.

Han (2012, p. 7) explains that in South Korea there are already over 1500 robots in use in preschool contexts and over 30 English education robots currently in active use in elementary school after-school programs.

Some studies have been conducted on the use of robots for language learning, particularly with young children. Several studies have concluded that robots can increase motivation and enhance language learning (e.g., Han, Jo, Park, & Kim, 2005; Movellan, Eckhardt, Virnes, & Rodriguez, 2009; Park et al., 2011). Han et al. (2005) compared English language learners' achievements with a robot, *IROBI*, versus with a standard computer and found *IROBI* to be more effective. In a study of pre-school children over a four-week period, Hyun, Kim, Jang, and Park (2008) found the robot more effective than standard computers for story building, vocabulary, understanding, and word recognition in Korean reading activities. Little research has appeared to date into the effectiveness of robots specifically for pronunciation learning, although studies suggest that young learners particularly enjoy oral activities, such as choral chanting, face-to-face conversation, and role-play, with the robot as teaching assistant (Han & Kim, 2009). As with other areas of CAPT, if robots represent the next technological paradigm shift, educators will need to give a great deal of consideration as to how best to use and integrate such devices from the perspective of pronunciation learning and pedagogy.

MOOCs

If robotics and smart technologies represent technology-led potential for changes in education and language learning, MOOCs (Massive Open Online Courses) represent a shift in pedagogic direction. A MOOC may be a “taster” course offered by a university or by a company. They are typically free and although many thousands of people may register (the average enrolment in 2014 was 43,000 students), the dropout rate is often very high. They are usually content-driven, with short video clips of lectures followed by online quizzes and readings, and with social networking through online discussion boards, although there is rarely any direct communication between individual students and the instructor.

There are relatively few language learning MOOCs, which is not surprising given the nature of language learning as skill development and the difficulty of providing individualized feedback to thousands of participants. Godwin-Jones (2014b) suggests that “the optimal approach to structuring a language learning MOOC is to provide an adaptive learning system within an extensive social and personalizable learning environment” (p. 8). Godwin-Jones (2014b) describes several language learning MOOCs (such as *Arabic for Global Exchange*⁵¹ and *American English Speech*⁵²), which have used this model, incorporating an AI-based adaptive learning system together with social networking, using tools such as *Google Groups*⁵³ and *Google Hangouts*,⁵⁴ to provide opportunities for real language exchange. There are a few MOOCs focusing specifically on pronunciation, such as the Japanese *Pronunciation for Communication MOOC*⁵⁵ (Waseda University). However, most pronunciation-focused MOOCs are predominantly perception-based, with little evidence of opportunities for speech input or evaluation, despite claims such as “We are also going to study Chinese Mandarin phonetics” or “This is an elementary course on Chinese speaking” (*Chinese for Beginners MOOC*,⁵⁶ Beijing University).

It is likely that MOOCs are here to stay but their value for language learning, and particularly for developing specific skills like pronunciation, will depend on further advances and applications in AI and careful considerations of pedagogic utility and potentials.

Using Other Technological Resources for Pronunciation

As well as the many custom-built CAPT resources available, there are other technological applications which, while not aimed specifically at teaching or learning pronunciation, can be used to develop various pronunciation skills. Such resources can be used for targeted pronunciation practice, focusing for instance on perception or production and on specific subskills, such as accuracy, fluency, or impact (see Table 5.1). The internet can provide limitless access to a wide array of natural, diverse, and motivating content which can be harnessed for work on pronunciation for perception (e.g., using short sections of online TV news for dictation or transcription, focusing on accuracy) or production (e.g., shadowing a piece of dramatic dialogue from a film or recording a voice commentary for part of a documentary program, focusing on fluency and impact). These resources can be married to other technological resources such as smartphones and iPods in creative ways that can be geared to pronunciation. For example, we note how Foote and McDonough (2017) made use of the iPod with short recordings for students to practice shadowing as an aid to pronunciation. In a more elaborate instructional design using several different technologies, Fouz-González (2015a) paired podcasts of “BBC 6 minute English” with PowerPoint presentations of pronunciation features and tips for correcting mispronunciations delivered in an online platform (*Edmodo*⁵⁷) for sharing materials and for grouping and interacting with students, together with *Google Drive*’s online questionnaire utility. The latter was used for gathering anonymous data in weekly peer and self-evaluations of student recordings in which they attempted to reproduce segments of the podcasts focused on their pronunciation. Two groups received input focused on different aspects of form (either /s, z/ or /b, d, g/), with each group improving in some aspects of the forms they practiced and all finding the approach to working on pronunciation helpful for improving their perception and production of English sounds.

Technologies designed for other purposes are increasingly recognized for their potential to aid language learning and pronunciation. For example, a program called *The French Digital Kitchen*⁵⁸ (Seedhouse, 2017), originally developed as an assistive technology (see Chap. 7), was later recognized to have application to language learning, and its use of speech synthesis could have application to training pronunciation in particular.

A range of non-CAPT software can be used for pronunciation practice. For instance, dictation exercises can be carried out online via the *Dragon* software or other programs such as *Evernote*,⁵⁹ *TalkTyper*,⁶⁰ *VoiceAssistant*,⁶¹ *Speechlogger*,⁶² or *PaperPort*.⁶³ *Dragon Naturally Speaking*, the improved version of the original *Dragon Dictate*, is the leading app for utilizing the transcription of voice to text on a smartphone. One of the most convenient features of the app is its ability to learn phrases said often and so improve its functionality. The *Dragon* app transcribes the best guess as to what the user, such as an L2 learner, actually said. When the *Dragon* L2 transcription contains errors, the learner knows that the pronunciation has deviated from the statistical norms programmed into this app. In turn, this obliges the user to analyze and restate the utterance in a more comprehensible way, thus providing a feedback loop along the lines of “forced output” as described by Swain (2000) and Swain and Lapkin (1998). When combined with explicit instruction on L2 sound and grapheme equivalencies, such dictation activities can provide helpful practice for learners. Relatively few CALL studies or practical applications have taken advantage of dictation software, despite the obvious affordances for enhancing L2 speaking.

Voice-activated personal assistant apps for computer and mobile phone, such as Amazon *Alexa*, Apple *Siri*, and *Google Assistant*, offer potential for authentic, task-based, oral language practice. The voice commands and voice searching functions, which enable users to, for instance, send messages, check appointments, or find information, lend themselves to pair or group-based language activities such as organizing a trip to a theater or restaurant. Such real-world apps incorporating ASR technology may serve as a value-added learning factor motivating students to improve their pronunciation in order to fulfill a task. The limitations of ASR can be seen a challenge to students to make their pronunciation sufficiently clear as to be comprehensible to the software.

Translation apps such as Android’s *Google Translate* use voice to translate between many different languages. This kind of resource can be used to focus on specific skills; for example, the *Google Translate* app could be used for accuracy practice of minimal pairs or short phrases to translate from the target language into the learner’s L1. The focus could be placed on impact, by getting learners to prepare and record part of a play or poem, which they then upload as a video email to their fellow learners or

teacher, using an app such as *Tokbox*,⁶⁴ which enables the integration of live video, voice, and messaging into websites and mobile apps.

Social networking media, such as *Skype*, *Twitter*, or *Facebook*, can also be used for pronunciation practice. Mompean and Fouz-González (2016) explore the use of *Twitter* by students at a language school in Spain to practice the pronunciation of difficult words, finding that the use of daily tweets had a beneficial effect on the students' pronunciation of the target words. A more extensive investigation of *Twitter* for improving pronunciation was carried out by Fouz-González (2017⁶⁵) with 121 participants of an ESP course focused on English for Medicine using a comparative pretest/posttest experimental/control group design that included a second, delayed posttest. Participants who received daily tweets providing short explanations of pronunciation features and links to video or audio files illustrating the pronunciation of problematic words in context improved in their pronunciation significantly as compared to those in the control group. They also maintained their gains in pronunciation over the next month, up to the time the delayed posttest was administered. In addition, there was a modest tendency for those who showed greater engagement with the instructional treatment to improve more than those who were less engaged. Questionnaire findings about the intervention indicated that participants considered the input they received to be helpful and the use of *Twitter* to have good educational potential.

Some *Twitter* accounts directly focus on pronunciation, such as *Forvo*⁶⁶ (Pronunciation) which gives the pronunciation of English words and sentences spoken by a wide range of speakers from different countries and language backgrounds. Similarly, *Twitter* accounts like *Oxford Words*⁶⁷ or *Cambridge Words*⁶⁸ tweet "words of the day" and include a link to the online dictionaries where readers can find the definition and pronunciation of the word. Some of these resources can be helpful for individual or self-study, for instance, using an online talking dictionary to quickly check the pronunciation of a word. Many tools, such as video email and voiceboards (i.e., audio discussion boards), lend themselves well to pair or group activities and to more extended project-based oral language practice while maintaining a focus on pronunciation.

A summary of some non-CAPT resources and their potential application for pronunciation skills training is given in Table 5.1.

Table 5.1 Some skill-focused non-CAPT resources

Activity focus	Skill focus	Example activity	Technology tools
Phoneme/syllable/ word stress recognition/ production	Accuracy	<i>Dictionary search</i> Students look up word meaning and listen to pronunciation using website or app or twitter. Students use app to look up loan words to compare pronunciation and word stress (e.g., "restaurant" in French and English).	Audio dictionary websites, apps, or <i>Twitter</i> (e.g., <i>Oxford English Dictionary</i> ; <i>Forvo</i>) Translation app (e.g., <i>Google Translate</i>)
Phoneme/minimal pairs practice	Accuracy	<i>Rhyming activities</i> Students use interactive rhyming game to predict/check minimal pairs.	Interactive educational websites (e.g., <i>Rhyme Zone</i> ⁶⁹ ; interactive sites for education)
Sound discrimination/ segmenting speech	Accuracy	<i>Video email</i> The teacher dictates a text and asks students to watch and write the text in the email and send it back to teacher for correction. <i>"Chinese (video) whispers"</i> The teacher records a short text and sends it by video email to one student who then re-records and forwards it to the next student until the final student sends the video message back to the teacher. The original and final versions are then compared (this idea from Peachey, 2009).	Video email app (e.g., <i>Tokbox</i> ; <i>Mailvu</i> ⁷⁰)
Noticing/ correcting pronunciation errors	Accuracy	<i>Audio discussion board</i> Students upload and share audio recordings on voiceboard for group discussion and peer/teacher feedback.	Voiceboards, i.e., e-audio discussion tools (e.g., <i>Voxopop</i> ⁷¹ ; <i>Voicethread</i> ⁷² ; <i>Voki</i> ⁷³)

(continued)

Table 5.1 (continued)

Activity focus	Skill focus	Example activity	Technology tools
Understanding accents/ connected speech	Fluency	<i>Video subtitling</i> A short video extract is transcribed by students in groups, to produce subtitles.	Online video site (e.g., <i>YouTube</i> ; <i>Vimeo</i>) Subtitling app (e.g., <i>Dotsub</i> ⁷⁴ ; <i>Fotobabble</i> ⁷⁵)
Recognizing/ producing connected speech	Fluency	<i>Video voice-over</i> A short video extract (e.g., from online news channel) is transcribed and the script is uploaded to a teleprompter app. Students then re-record the extract using the teleprompter and their phone video recorder or webcam.	<i>Screen capture app</i> (e.g., <i>Screencast-O-Matic</i> ⁷⁶) Teleprompter app (e.g., <i>Simple Teleprompter</i> ⁷⁷)
Producing connected speech (rhythm and stress)	Fluency	<i>Recording poems</i> Students download a poem from website, check pronunciation of new words with online dictionary and record a version using online recording software (idea from Peachey, 2009).	Poetry website (e.g., <i>HelloPoetry</i> ⁷⁸) Online talking dictionary (e.g., <i>Howjsay</i> ⁷⁹) Audio recording software (e.g., <i>Audacity</i> ⁸⁰)
Producing connected speech (rhythm and stress)	Fluency	<i>Karaoke</i> Students record their own song using online Karaoke site and then compare this with the original singer's version.	Karaoke website (e.g., <i>Karoke Party</i> ⁸¹ ; <i>Lucky Voice Karaoke</i> ⁸²)
Producing connected speech	Accuracy and fluency	<i>Planning a visit to a restaurant</i> Students (preferably working in pairs or a group) use voice-activated app to find information, send messages and book a table at a restaurant.	Voice-activated personal assistant apps (e.g., <i>Amazon Alexa</i> ⁸³ ; <i>Apple Siri</i> ⁸⁴ ; <i>Microsoft Cortana</i> ⁸⁵)

(continued)

Table 5.1 (continued)

Activity focus	Skill focus	Example activity	Technology tools
Modelling intonation patterns	Impact	<i>Reading a film script</i> Students find film script from movie website and select short dialogue or monologue to model. They mark intonation (pauses, stress placement and pitch movement) and then. Reproduce extract.	Movie script website (e.g., <i>IMSDB</i> , <i>Internet Movie Script database</i> ⁸⁶)
Shadowing intonation patterns	Impact	<i>Shadowing dramatic or powerful speech</i> Students transcribe a short extract from famous video speech and mark intonation (pauses, stress placement and pitch movement). They then practice shadowing the extract (i.e., reading it aloud at the same time as the speaker).	You Tube video (e.g., Martin Luther King's "I have a dream" speech ⁸⁷)

Effectiveness of CAPT

One of the key concerns regarding computer-based pronunciation learning, as with pronunciation learning and teaching in general, is whether or not it is effective. Lee, Jang, and Plonsky's (2015) meta-analysis of pronunciation teaching effectiveness studies concludes that research into the use of technology in pronunciation teaching showed an overall smaller effect than human-based pronunciation teaching, although Lee et al. (2015) observe that their negative finding for computer-based pronunciation teaching may be due to "[t]he lack of adaptability and perceptual accuracy in computers compared to human teachers, and perhaps consequently their ability to provide appropriate feedback as well" (p. 360). However, as we have seen, there is some evidence that CAPT can be effective in pronunciation learning, particularly for perceptual training and improvement of prosody through visual feedback. Effective automated

feedback may still be some way off, though convincing benefits have been shown using speech analysis software with the guidance of a teacher (Hardison, 2004; Pearson, Pickering, & Da Silva, 2011). Research studies consistently show that multimodal input can be beneficial to learning, and this is clearly an area where CAPT has an advantage, for instance, being able to provide 3D computer animations of the lips and oral cavity and interior vocal structures together with audio to help with the articulation of sounds (Albertson, 1982; Ehsani & Knodt, 1998; Elliot, 1995; Grant & Greenberg, 2001).

Studies suggest that pronunciation accuracy can be improved in a number of dimensions with well-designed CAPT resources. At the segmental level, for example, Kawai and Hirose (2000) found that Japanese learners were able to more successfully distinguish phonemic contrasts between English long and short vowels using speech recognition software. Wang and Munro (2004) documented an improvement in learners' perception of a number of English vowel contrasts (/i/-/ɪ/, /u/-/ʊ/, and /ɛ/-/æ/) after using training software based on perceptual identification tasks with synthetic and natural speech in various voices. Hirata (2004) reported improved production of pitch and duration contrasts by learners of Japanese using a pronunciation training program that provided fundamental frequency contours as visual feedback. At the suprasegmental level, various studies have shown that CAPT programs can help learners with the perception and production of intonation, especially when visual representations of learners' prosody are provided as feedback in comparison to a model contour (e.g., de Bot & Mailfert, 1982; Hardison, 2004; Levis & Pickering, 2004). There is also some evidence to show that CAPT can be beneficial for children, if carefully designed. As Mich et al. (2006) found for young foreign language learners, "training with a computer-assisted pronunciation training system with a simple automatic speech recognition component can...lead to short-term improvements in pronunciation that are comparable to those achieved by means of more traditional, teacher-led pronunciation training" (p. 393). As with other areas of pronunciation research, however, there is a general lack of studies which have considered long-term pronunciation gains based on any kind of computer-based input or practice.

Research shows that personalized, immediate feedback is particularly effective for language learning. Pronunciation feedback is an area that is still challenging CAPT developers, especially for feedback on learner performance using ASR. In the view of Neri et al. (2002), “ideal systems should always include an option to provide feedback by means of ASR technology, so that the user can receive immediate information on his/her performance” (p. 458). Ideal systems should be able to provide accurate feedback on prosodic as well as segmental features of speech and should prioritize significant errors in terms of frequency and perceptual relevance.

A further challenge with effectiveness studies in CAPT is that many resources aim to help learners develop native-like pronunciation, rather than focusing on intelligibility or other aspects of pronunciation competence that can be defined without reference to a native-speaker criterion or scale for performance. At present, technology is largely limited to comparing or matching learner performance with a database of native speaker models. Again, this is partly due to current limitations of ASR technology in not being able to recognize multiple models in multiple language varieties or accents. Research has been ongoing to enable effective multilingual speech processing (e.g., Goronzy, Rapp, & Kompe, 2004; Oh, Yoon, & Kim, 2007; Schultz & Kirchoff, 2006), but much of this research is currently of little direct relevance to CAPT. Some advances have been made towards producing reliable ASR feedback for L2 learners, such as Cucchiarini et al.’s (2009) CAPT system using ASR feedback for learners of Dutch. Much remains to be done to improve the technological capabilities of hardware and software for pronunciation training, but at the same time much more creative thought and design work needs to go into development of pronunciation pedagogy in terms of both the content and the methodology of instruction.

Computer-Based Pronunciation Assessment

There are considerable potential benefits in the use of computers for language testing or assessment, including the removal of human bias, for example, due to accent familiarity or preference, or fluctuations in

levels of concentration or tolerance (see Chap. 6). Theoretically, ASR is promising for pronunciation assessment; indeed, as Neumeyer, Franco, Digalakis, and Weintraub (2000) claim, “speech recognition technology is key to the automatic evaluation of pronunciation quality” (p. 83). ASR can easily detect deviations from model or standard speech/pronunciation on a global level numerical score, even if it cannot provide detailed feedback on the cause or significance of individual errors. However, the limited accuracy of ASR, as described earlier, means that there are considerable risks attached to computer-based assessment, particularly for high-stakes tests where a poor score may dash career hopes or destroy educational dreams. As Goronzy, Tomokiyo, Barnard, and Davel (2006) point out: “To date automatic speech recognition has not advanced to the point where it can be an alternative for human scorers on important evaluations of spoken language proficiency” (p. 309).

The ability to correlate ASR judgements of pronunciation quality against human judgements remains a key challenge. In one study, Neri et al. (2002) found that automated scoring detected only 25% of pronunciation errors and marked some accurate pronunciation features as errors. A study by Cucchiari, Strik and Boves (2000a) comparing human and machine scores of accented Dutch found that ASR-based scoring of segmental quality did not correlate well with expert human raters’ scores: “overall pronunciation is most influenced by segmental quality, which is the human measure that can be predicted most poorly on the basis of our machine scores” (p. 118). However, their study found a close correlation between human ratings and machine-generated measures of speech rate and duration that matches the results of early work by Pennington (1990, 1992) and has since been confirmed in other studies (e.g., Kang, Rubin, & Pickering, 2010; Kormos & Dénes, 2004). Cucchiari et al. (2000a) concluded that further work is needed to develop an automatic pronunciation test of Dutch and that “finding an adequate automatic correlate of segmental quality is necessary to avoid that fast speakers with low proficiency get high pronunciation scores” (p. 119). While we agree with this point, we also note the mounting evidence that rate measures, especially those that factor in the frequency and distribution of pauses in spontaneous speech (Cucchiari, Strik, Binnenpoorte, & Boves, 2002), are highly correlated with, and predictive

of, high ratings in L2 pronunciation and overall speaking proficiency. These higher scores presumably represent more rapid and fluent, automatized or routinized, cognitive processing and articulatory production (Chap. 1), resulting from increased L2 knowledge and experience speaking and requiring less explicit attention (Chap. 2).

It appears that the different measures used for automated scoring, combined with the variable content or quality of the speech model database to which non-native learner speech is compared, causes considerable variation in the performance of automated scoring systems. One of the key challenges is defining the appropriate criteria for the computer to use as the basis of measuring pronunciation proficiency, which relates closely to human ratings of proficiency. A fundamental issue is that machines cannot hear or process speech like humans and are not as good at dealing with the variations and complexities of natural speech. According to Scharenborg (2007), humans are “far better at dealing with accents, noisy environments, differences in speaking style, speaking rate, etc.” (p. 344). It would seem then that considerably more work is needed to clarify the parameters needed by automated ASR systems to detect pronunciation errors or deviations from speech models, particularly at the segmental level. Nevertheless, it has been suggested (Levis, 2007) that machine rating may be good enough at providing global pronunciation scores to be used in some types of automatic pronunciation evaluation. A case in point is the introduction of fully computerized speaking tests such as Pearson’s *Versant* tests and TOEFL’s *SpeechRater*, which is used to score the iBT Practice Online test (see Chap. 6).

The commercially successful *Versant*⁸⁸ (formerly *PhonePass*) test (Bernstein, Van Moere, & Cheng, 2010) uses ASR to evaluate the correctness of student responses (in the sense that it analyzes the match of the user’s input to that of its databases) and also gives scores for pronunciation and fluency. The *Versant* system uses ASR technology and acoustic measures of the properties of speech waves to identify units of speech and other computer programs to determine their characteristics for assessment purposes and to assign specific scores. For the measurement of pronunciation and fluency, the programs are able to identify specific words, the consonant and vowel phonemes within these, gaps in speech, and the timing of speech output.

Versant is often used for high-stakes purposes, for instance, in the aviation industry, where it is used to certify the language proficiency level of pilots and air traffic controllers. Pearson Education Inc. (2011) reports that the English speaking test has been extensively field trialed and validated in relation to human ratings of language proficiency, fluency, and pronunciation; and there is research specifically on its use with pilots and air traffic controllers (Balogh, Bernstein, Suzuki, & Lennig, 2011; Hincks, 2001) independently examined the *PhonePass* scores of a group of students and found a relationship between the students' rates of speech and their overall scores when reading the test sentences. We note that this finding is not necessarily negative for *PhonePass* but can be interpreted as a potential validation measure for this electronic testing system, given the predictive relationship found between measures of speech rate and human global or composite pronunciation or oral proficiency scores (e.g., Cucchiariini et al., 2000a, 2002, Cucchiariini, Strik, & Boves, 2000b; Kang et al., 2010; Kormos & Dénes, 2004; Pennington, 1990, 1992). Hincks (2001) highlighted some potential problems with the *PhonePass* system, including the absence of a way to measure the effects of prosody,⁸⁹ and the limitations of assessing only short samples of speech. Since the time of Hincks' (2001) study, *PhonePass* has been further developed in the *Versant* system. Nonetheless, the question remains as to whether the value of such computerized tests is restricted to highly predictable L2 speaking tasks or can be extended to evaluating discourse-level speech, which is the aim of the *SpeechRater* system in development by TOEFL (for further discussion, see Chap. 6).

Concluding Remarks

Technology holds great potential for pronunciation training, particularly in terms of maximizing opportunities for practice and exposure to spoken language, yet it has a way to go before it can be recommended as a totally stand-alone method of language learning. Some of the main criticisms of CAPT reflect a general disappointment that the field has not yet fulfilled many enthusiasts' earlier expectations and predictions (Derwing & Munro, 2015; Levis, 2007; Pennington, 1999). Some of these short-

comings are due to technological difficulties, as explained earlier, for instance, with improving ASR so that it handles prosody and non-native speech more accurately than at present. Another important issue is the continuing lack of connection between many commercially developed CAPT resources and research and theory about L2 phonological acquisition and teaching. This means that there is often a gap between learner needs and product features and that an authoritative basis for teacher decision-making regarding many CAPT resources is lacking.

The gap between learner needs and CAPT products could be lessened with closer collaborations between phonologists, teaching experts, and technical developers to create programs with clear pedagogic aims based on sound methodological principles and practices. Such collaboration could result in products in which functionalities are selected on the basis of pedagogic value as well as technical capabilities. Collaborations could also help ensure the quality of resources in terms of the validity and accuracy of the learning content and the reliability of feedback. Reliability and accuracy of feedback are crucial in language learning; nothing is more frustrating for learners than a system that provides different responses to successive realizations of the same input. Given the current technological limitations of providing specific feedback or diagnosis of a learner's pronunciation errors or problems, it may be necessary to compromise in CAPT regarding the amount of detail offered on errors and on the accuracy of feedback. As Neri et al. (2002) state, "... if we want to reach an ideal compromise between technology and demand, we will have to settle for something that is less ambitious, but that can guarantee correct feedback at least in the majority of the cases" (p. 459).

According to Neri et al. (2002), CAPT resources should provide opportunities for speech input and output, preferably incorporating ASR technology, and for customized, accurate feedback. Although entirely effective automated feedback remains an elusive goal, the benefits of new technologies could be more fully embraced and maximized. For instance, the potential has hardly been explored yet for incorporating multimedia through film extracts, radio interviews, and podcasts with a wide range of L1 and L2 speaker accents, using 3D animations of speech articulations, or exploiting social media such as Twitter for pronunciation learning (Fouz-González, 2015a, 2017; Mompean & Fouz-González, 2016).

Similarly, adopting a games-based approach with simulations and role-plays would put more emphasis on realistic, task-based pronunciation learning activities in CAPT. As with other areas of language instruction, teachers need clear guidance and more knowledge of the benefits and uses of CAPT, so that they can make informed, critical choices about which resources are most useful for their learners and can integrate technologies meaningfully into language teaching. Also, more research is needed on the effectiveness of CAPT, particularly, the longer-term benefits, as well as more detailed, critical reviews of popular commercial software, such as the review by Lin, Warschauer, and Blake (2016) of *Livemocha* and Lord's (2015) review of *Rosetta Stone*. Despite some of the obstacles mentioned here, as technologies continue to advance, their applications for pronunciation teaching and learning will improve. In the meantime, CAPT offers the affordances of multimodality, mobility, autonomy, as well as endless opportunities for practice and exposure to a wide variety of speech.

Notes

1. <https://wavesurfer-js.org/>
2. <http://www.fon.hum.uva.nl/praat/>
3. <https://www.nuance.com/en-gb/dragon.html>
4. <https://www.apple.com/uk/ios/siri/>
5. <https://www.amazon.com/Amazon-Echo-And-Alexa-Devices/?ie=UTF8&node=9818047011>
6. <https://www.microsoft.com/en-gb/windows/cortana>
7. <http://www.englishlearning.com/>
8. <http://www.macmillaneducationapps.com/soundspron/>
9. <https://www.eda.kent.ac.uk/medical/linguagraph.aspx>
10. <https://www.eda.kent.ac.uk/medical/linguaview.aspx>
11. <https://www.proteatextware.com/shopexd.asp?id=66&bc=no>
12. <http://www.speechinaction.org/cool-speech-app-banner/>
13. <http://www.englishaccentcoach.com/>
14. <http://soundsofspeech.uiowa.edu/>
15. https://www.llas.ac.uk/sites/default/files/mbfiles/mb081/page_01.htm
16. <http://www.phon.ucl.ac.uk/resource/tutorials.html#phon>

17. <http://davidbrett.it/index.php?home>
18. <http://www.tedpower.co.uk/phono.html>
19. <https://www.duolingo.com/>
20. <https://www.pentaxmedical.com/pentax/en/94/1/Visi-Pitch-IV-Model-3950B-Computerized-Speech-Lab-CSL-Model-4500-and-4150B>
21. <http://www.winpitch.com/>
22. <http://www.betteraccent.com/>
23. <https://play.google.com/store/apps/details?id=com.androidrockervoicechanger&hl=en>
24. <https://www.nuance.com/en-gb/dragon.html>
25. <http://www.englishlearning.com/products/pronunciation-power-speech-test/>
26. <https://www.tellmemorecampus.com/>
27. <https://www.englishlearning.com/products/pronunciation-power-speech-test>
28. <https://auralog.com>
29. <https://www.effectivelanguagelearning.com/language-course-reviews/tell-me-more-review>
30. <https://learnenglishkids.britishcouncil.org/en/helping-your-child/pronunciation-activities>
31. http://www.esl.net/tell_me_more_kids.html
32. <https://itunes.apple.com/us/app/phonics-studio/id547795266?mt=8>
33. <https://www.commonsemmedia.org/app-reviews/c-is-for-cow>
34. http://littlebeespeech.com/articulation_station.php
35. https://elt.oup.com/catalogue/items/global/adult_courses/english_file_third_edition/9780194597982?cc=us&sellLanguage=en
36. <http://www.macmillaneducationapps.com/soundspron/>
37. [https://play.google.com/store/apps/developer?id=Cambridge+Learning+\(Cambridge+University+Press\)&hl=en](https://play.google.com/store/apps/developer?id=Cambridge+Learning+(Cambridge+University+Press)&hl=en)
38. The iPad app was “retired” and removed from the Apple Store on May 3, 2017, and it does not work with the latest iOS on the iPad. As reported on the Speech in Action website (<http://www.speechinaction.org/projects/cool-speech-ipad-app-update/>), “the level of sales does not warrant spending money on an upgrade.”
39. <https://www.duolingo.com/>
40. <https://www.elsaspeak.com/home>
41. <http://supiki.com/>
42. <https://www.jabberwacky.com>
43. <https://www.mondlylanguages.com/>

44. <https://www.mondly.com/blog/2016/08/25/mondly-chatbot-press-release/>
45. <http://psyentificmind.com/language-instruction-with-baldi>
46. <http://www.speech.kth.se/ville/>
47. <http://www.speech.kth.se/deal/>
48. <https://learning2gether.net/tag/traci-talk/>
49. <https://gamerlearner.com/2007/12/07/croquelandia-uni-of-minnesota-spanish-language-virtual-world-project/>
50. <https://www.alelo.com/case-study/tactical-iraqi-language-culture-training-system/>
51. <http://oli.cmu.edu/courses/all-oli-courses/arabic-for-global-exchange/>
52. <http://oli.cmu.edu/courses/all-oli-courses/speech-course-details/>
53. <https://groups.google.com/forum/#!overview>
54. <https://hangouts.google.com/>
55. <https://www.edx.org/course/japanese-pronunciation-communication-wasedax-jpc111x-1>
56. <https://www.coursera.org/learn/learn-chinese>
57. <https://www.edmodo.com/>
58. https://europeandigitalkitchen.com/?page_id=402
59. <https://evernote.com/>
60. <https://talktyper.com/>
61. <https://itunes.apple.com/us/app/voice-assistant-just-use-your-voice-instead-of-typing/id511757903?mt=8>
62. <https://speechlogger.appspot.com/en/>
63. <https://www.nuance.com/en-gb/print-capture-and-pdf-solutions/optical-character-recognition/paperport-for-pc.html>
64. <https://tokbox.com/>
65. This study was originally carried out as one of several investigations that were part of Fouz-González's (2015a) PhD thesis, for which Martha served as external examiner.
66. <https://forvo.com/>
67. <https://blog.oxforddictionaries.com/>
68. <https://dictionaryblog.cambridge.org/author/cambridgewords/>
69. <http://interactivesites.weebly.com/rhyming.html>
70. <http://mailvu.com/>
71. <https://www.voxpop.com/>
72. <https://voicethread.com/>
73. <http://www.voki.com/>
74. <https://dotsub.com/>
75. <http://www.fotobabble.com/>

76. <https://screencast-o-matic.com/>
77. https://play.google.com/store/apps/details?id=com.schakib.julian.teleprompter&hl=en_GB
78. <https://helopoetry.com/>
79. <http://www.howjsay.com/?&wid=2552&flash=n>
80. <https://www.audacityteam.org/>
81. <https://www.redkaraoke.com/search-songs/Party-Songs/3>
82. <https://www.luckyvoicekaraoke.com/>
83. <https://www.amazon.com/Amazon-Echo-And-Alexa-Devices/b?ie=UTF8&node=9818047011>
84. <https://www.apple.com/uk/ios/siri/>
85. <https://www.microsoft.com/en-gb/windows/cortana>
86. <http://www.imsdb.com/>
87. <https://www.youtube.com/watch?v=smEqnklfYs>
88. <https://www.versanttests.com/>
89. Also found by Kang et al. (2010) to be an important predictor of human ratings of pronunciation.

References

- Albertson, K. (1982). Teaching pronunciation with visual feedback. *NALLD Journal*, 17, 18–33.
- Alsabaan, M., & Ramsay, A. (2014). Diagnostic CALL tool for Arabic learners. In S. Jager, L. Bradley, E. J. Meima, & S. Thouëсны (Eds.), *CALL design: Principles and practice, Proceedings of the 2014 EUROCALL Conference*, Groningen, The Netherlands (pp. 6–11). Dublin: Research-publishing.net. <https://doi.org/10.14705/rpnet.2014.000186>
- Altman, J. (2013). Taming the dragon: Effective use of dragon naturally speaking speech recognition software as an avenue to universal access. *Writing & Pedagogy*, 5(2), 333–348. <https://doi.org/10.1558/wap.v5i2.333>
- Balogh, J., Bernstein, J., Suzuki, M., & Lennig, M. (2011). *Automatically scored spoken language tests for air traffic controllers and pilots*. Paper presented at International Aviation Training Symposium 2006, Oklahoma City. Versant White Paper. Retrieved February 22, 2018, from http://www.versanttest.com/technology/featured/Developmental_Paper_AET_IATS2006-reformatted.pdf
- Benoît, C., Mohammadi, T., & Kandel, S. (1994). Effects of phonetic context on audio-visual intelligibility of French. *Journal of Speech and Hearing Research*, 37(5), 1195–1203. <https://doi.org/10.1044/jshr.3705.1195>

- Bernstein, J., Van Moere, A., & Cheng, J. (2010). Validating automated speaking tests. *Language Testing*, 27(3), 355–377. <https://doi.org/10.1177/0265532210364404>
- Bosseler, A., & Massaro, D. W. (2003). Development and evaluation of a computer-animated tutor for vocabulary and language learning for children with autism. *Journal of Autism and Developmental Disorders*, 33(6), 653–672. <https://doi.org/10.1023/B:JADD.0000006002.82367.4f>
- Chou, F. (2005). Ya-Ya language box: A portable device for English pronunciation training with speech recognition technologies. *Proceedings of Interspeech 2005*, Lisbon, Portugal (pp. 169–172). Retrieved February 22, 2018, from http://www.isca-speech.org/archive_papers/interspeech_2005/i05_0169.pdf
- Chun, D. M. (2013). Computer-assisted pronunciation teaching. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics*. Oxford: Wiley-Blackwell. <https://doi.org/10.1002/9781405198431.wbeal0172>
- Coniam, D. (1999). Voice recognition software accuracy with second language speakers of English. *System*, 27(1), 49–64. [https://doi.org/10.1016/S0346-251X\(98\)00049-9](https://doi.org/10.1016/S0346-251X(98)00049-9)
- 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. <https://doi.org/10.1080/09658410208667044>
- Cucchiari, C., Neri, A., & Strik, H. (2009). Oral proficiency training in Dutch L2: The contribution of ASR-based corrective feedback. *Speech Communication*, 51(10), 853–863. <https://doi.org/10.1016/j.specom.2009.03.003>
- Cucchiari, C., Strik, H., Binnenpoorte, D., & Boves, L. (2002). Pronunciation evaluation in read and spontaneous speech: A comparison between human ratings and automatic scores. In A. James & J. Leather (Eds.), *New sounds 2000, Proceedings of the Fourth International Symposium on the Acquisition of Second-Language Speech*, University of Amsterdam, September 2000 (pp. 72–79). Klagenfurt, Austria: University of Austria.
- Cucchiari, C., Strik, H., & Boves, L. (2000a). Different aspects of expert pronunciation quality ratings and their relation to scores produced by speech recognition algorithms. *Speech Communication*, 30(2-s3), 109–119. [https://doi.org/10.1016/S0167-6393\(99\)00040-0](https://doi.org/10.1016/S0167-6393(99)00040-0)
- Cucchiari, C., Strik, H., & Boves, L. (2000b). Quantitative assessment of second language learners' fluency by means of automatic speech recognition technology. *Journal of the Acoustical Society of America*, 107(2), 989–999. <https://doi.org/10.1121/1.428279>
- de Bot, K. (1983). Visual feedback of intonation I: Effectiveness and induced practice behavior. *Language and Speech*, 26(4), 331–350. <https://doi.org/10.1177/002383098302600402>

- de Bot, K., & Mailfert, K. (1982). The teaching of intonation: Fundamental research and classroom applications. *TESOL Quarterly*, 16(1), 71–77. <https://doi.org/10.2307/3586564>
- de Meo, A., Vitale, M., Pettorino, M., Cutugno, F., & Origlia, A. (2013). Imitation/self-imitation in computer assisted prosody training for Chinese learners of L2 Italian. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 4th pronunciation in second language learning and teaching conference*, August 2012 (pp. 90–100). Ames, IA: Iowa State University.
- Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. Amsterdam: John Benjamins.
- Derwing, T. M., Munro, M. J., & Carbonaro, M. D. (2000). Does popular speech recognition software work with ESL speech? *TESOL Quarterly*, 34(4), 592–603. <https://doi.org/10.2307/3587748>
- Díaz-Vera, J. (2012). Great expectations: Formalizing and transforming mobile-assisted language learning. In J. Díaz-Vera (Ed.), *Left to my own devices: Learner autonomy and mobile-assisted language learning* (pp. xi–xix). Bingley, UK; Leiden and Boston: Emerald; Brill.
- Ehsani, F., & Knodt, E. (1998). Speech technology in computer-aided language learning: Strengths and limitations of a new CALL paradigm. *Language Learning and Technology*, 2(1), 54–73. <http://dx.doi.org/10.125/25032>
- Elliot, R. (1995). Foreign language phonology: Field independence, attitude and the success of formal instruction in Spanish pronunciation. *Modern Language Journal*, 79(4), 530–542. <https://doi.org/10.1111/j.1540-4781.1995.tb05456.x>
- Eskenazi, M. (2009). An overview of spoken language technology for education. *Speech Communication*, 51(10), 832–844. <https://doi.org/10.1016/j.specom.2009.04.005>
- Felps, D., Bortfeld, H., & Gutierrez-Osuna, R. (2009). Foreign accent conversion in computer assisted pronunciation training. *Speech Communication*, 51(10), 920–932. <https://doi.org/10.1016/j.specom.2008.11.004>
- Foote, J. A., & McDonough, K. (2017). Using shadowing with mobile technology to improve L2 pronunciation. *Journal of Second Language Pronunciation*, 3(1), 34–56. <https://doi.org/10.1075/jslp.3.1.02foo>
- Fouz-González, J. (2012). Can Apple's iPhone help to improve English pronunciation autonomously? State of the app. In L. Bradley & S. Thouësný (Eds.), *CALL: Using, Learning, Knowing, EUROCALL Conference Proceedings* (pp. 81–87), Gothenburg, Sweden, 22–25 August 2012. Dublin, Ireland: Research-publishing.net
- Fouz-González, J. (2015a). *Foreign language pronunciation training with affordable and easily accessible technologies: Podcasts, smartphone apps and social networking services (Twitter)*. PhD thesis, University of Murcia (Spain).

- Fouz-González, J. (2015b). Trends and directions in computer assisted pronunciation training. In J. Mompean & J. Fouz-González (Eds.), *Investigating English pronunciation: Trends and directions* (pp. 314–342). Basingstoke, UK and New York: Palgrave Macmillan.
- Fouz-González, J. (2017). Pronunciation instruction through Twitter: The case of commonly mispronounced words. *Computer-Assisted Language Learning*, 30(7), 631–663. <https://doi.org/10.1080/09588221.2017.1340309>
- Fryer, L., & Carpenter, R. (2006). Emerging technologies: Bots as language learning tools. *Language Learning & Technology*, 10(3), 8–14. Retrieved January 1, 2018, from <http://llt.msu.edu/vol10num3/emerging/>
- Germain-Rutherford, A., & Martin, P. (2002). Integration of speech technology in oral expression courses: Focus on feedback. In *Proceedings of the conference on Technology in Language Education: Meeting the Challenges of Research and Practice* (pp. 39–44). Hong Kong and Nanjing. Retrieved August 17, 2017, from https://www.researchgate.net/publication/241758693_Integration_of_speech_technology_in_oral_expression_courses_Focus_on_feedback
- Gerosa, M., & Giuliani, D. (2004). Preliminary investigations in automatic recognition of English sentences uttered by Italian children. In R. Delmonte, P. Delcloque, & S. Tonelli (Eds.), *Proceedings of NLP and Speech Technologies in Advanced Language Learning Systems Symposium*, Venice, Italy (pp. 9–12). Padova: Unipress.
- Gibbon, F. & Lee, A. (2007). Electropalatography as a research and clinical tool. In L. Vallino-Napoli (Ed.), *Perspectives on Speech Science and Orofacial Disorders* (ASHA Division 5), 17, 7–13.
- Gilbert, J. B. (2012). *Clear speech: Pronunciation and listening comprehension in North American English* (4th ed.). New York: Cambridge University Press.
- Godwin-Jones, R. (2014a). Games in language learning: Opportunities and challenges. *Language Learning & Technology*, 18(2), 9–19. Retrieved January 1, 2018, from <http://llt.msu.edu/issues/june2014/emerging.pdf>
- Godwin-Jones, R. (2014b). Global reach and local practice: The promise of MOOCS. *Language Learning & Technology*, 18(3), 5–15. Retrieved January 1, 2018, from <http://llt.msu.edu/issues/october2014/emerging.pdf>
- Golonka, E. M., Bowles, A. R., Frank, V. M., Richardson, D. L., & Freynik, S. (2014). Technologies for foreign language learning: A review of technology types and their effectiveness. *Computer Assisted Language Learning*, 27(1), 70–105. <https://doi.org/10.1080/09588221.2012.700315>
- Goronzy, S., Rapp, S., & Kompe, R. (2004). Generating non-native pronunciation variants for lexicon adaptation. *Speech Communication*, 42(1), 109–123. <https://doi.org/10.1016/j.specom.2003.09.003>

- Goronzy, S., Tomokiyo, L. M., Barnard, E., & Davel, M. (2006). Other challenges: Non-native speech, dialects, accents, and local interfaces. In T. Schultz & K. Kirchhoff (Eds.), *Multilingual speech processing* (pp. 273–316). San Diego and London: Academic Press.
- Grant, K., & Greenberg, S. (2001). *Speech intelligibility derived from asynchronous processing of auditory visual information*. Paper presented at the Audio-visual Speech Processing Workshop 2001. Retrieved February 22, 2018, from <http://www.icsi.berkeley.edu/ftp/pub/speech/papers/avsp01-av.pdf>
- Hacker, C., Batliner, A., Steidl, S., Nöth, E., Niemann, H., & Cincarek, T. (2005). Assessment of non-native children's pronunciation: Human marking and automatic scoring. In G. Kokkinakis, N. Fakotakis, & E. Dermatas (Eds.), *Proceedings of the 10th International Conference on Speech and Computer (SPECOM 2005)*, Patras, Greece, Vol. 1 (pp. 123–126). Moscow: Moscow State Linguistic University.
- Han, J. (2012). Emerging technologies: Robot-assisted language learning. *Language Learning & Technology*, 16(3), 1–9. Retrieved January 1, 2018, from <http://llt.msu.edu/issues/october2012/emerging.pdf>
- Han, J., Jo, M., Park, S., & Kim, S. (2005). The educational use of home robots for children. In *Proceedings of the 14th IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN 2005)*, Nashville, TN (pp. 378–383). Piscataway, NJ: IEEE. <https://doi.org/10.1109/ROMAN.2005.1513808>
- Han, J., & Kim, D. (2009). r-Learning services for elementary school students with a teaching assistant robot. In *Proceedings of the 4th ACM/IEEE Human Robot Interaction*, La Jolla, CA (pp. 255–256). New York: ACM. <https://doi.org/10.1145/1514095.1514163>
- Harashima, H. D. (1999). Software review: *Tracy [sic] Talk—The Mystery. Computer Assisted Language Learning*, 12(3), 271–274. <https://doi.org/10.1076/call.12.3.271.5708>
- Hardison, D. M. (2004). Generalization of computer-assisted prosody training: Quantitative and qualitative findings. *Language Learning and Technology*, 8(1), 34–52. Retrieved January 1, 2018, from <http://llt.msu.edu/vol8num1/Hardison>
- Hardison, D. M. (2007). The visual element in phonological perception and learning. In M. C. Pennington (Ed.), *Phonology in context* (pp. 135–158). New York: Palgrave Macmillan.
- Hill, C. J. (2013). Apple's dictation software: A voice solution for writers whose hands need a rest. *Writing & Pedagogy*, 5(2), 346–355. <https://doi.org/10.1558/wap.v5i2.349>
- Hincks, R. (2001). Using speech recognition to evaluate skills in spoken English. Lund University, Dept. of Linguistics *Working Papers*, 49, 58–61. Retrieved

- September 23, 2017, from journals.lub.lu.se/index.php/LWPL/article/download/2367/1942/
- Hincks, R. (2005). Measures and perceptions of liveliness in student oral presentation speech: A proposal for automatic feedback mechanism. *System*, 33(4), 575–591. <https://doi.org/10.1016/j.system.2005.04.002>
- Hincks, R., & Edlund, J. (2009). Promoting increased pitch variation in oral presentations with transient visual feedback. *Language Learning and Technology*, 13(3), 32–50 <https://dx.doi.org/10.125/44190>
- Hirata, Y. (2004). Computer-assisted pronunciation training for native English speakers learning Japanese pitch and duration contrasts. *Computer Assisted Language Learning*, 17(3–4), 357–376. <https://doi.org/10.1080/0958822042000319629>
- Holland, V., Kaplan, J., & Sabol, M. (1999). Preliminary tests of language learning in a speech-interactive graphics microworld. *CALICO Journal*, 16(3), 339–359. <https://doi.org/10.1558/cj.v16i3.339-359>
- Hyun, E.-j., Kim, S.-y., Jang, S., & Park, S. (2008). Comparative study of effects of language education program using intelligence robot and multimedia on linguistic ability of young children. In *Proceedings of the 17th IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN 2008)*, Munich, Germany (pp. 87–192). Piscataway, NJ: IEEE. <https://doi.org/10.1109/ROMAN.2008.4600664>
- Jesse, A., Vrignaud, N., Cohen, M. M., & Massaro, D. W. (2000). The processing of information from multiple sources in simultaneous interpreting. *Interpreting*, 5(2), 95–115. <https://doi.org/10.1075/intp.5.2.04jes>
- Kachru, B. B. (1985). Standards, codification and sociolinguistic realism: The English language in the outer circle. In R. Quirk & H. Widdowson (Eds.), *English in the world: Teaching and learning the language and literatures* (pp. 11–30). Cambridge: Cambridge University Press.
- Kaltenboeck, G. (2002). Computer-based intonation teaching: Problems and potential. In *Talking Computers, Proceedings of the IATEFL Pronunciation and Computer Special Interest Groups* (pp. 11–17). Whitstable, UK: IATEFL.
- Kang, O., Rubin, D., & Pickering, L. (2010). Suprasegmental measures of accentedness and judgments of language learner proficiency in oral English. *The Modern Language Journal*, 94(4), 554–566. <https://doi.org/10.1111/j.1540-4781.2010.01091.x>
- Kawai, G., & Hirose, K. (2000). Teaching the pronunciation of Japanese double-mora phonemes using speech recognition technology. *Speech Communication*, 30(2–3), 131–143. [https://doi.org/10.1016/S0167-6393\(99\)00041-2](https://doi.org/10.1016/S0167-6393(99)00041-2)

- Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, 32(2), 145–164. <https://doi.org/10.1016/j.system.2004.01.001>
- Krajka, J. (2001). *English for Kids*. CALICO Software Review. Retrieved January 1, 2018, <https://journals.equinoxpub.com/index.php/CALICO/article/view/35173/pdf>
- Lee, J., Jang, J., & Plonsky, L. (2015). The effectiveness of second language pronunciation instruction: A meta-analysis. *Applied Linguistics*, 36(3), 345–366. <https://doi.org/10.1093/applin/amu040>
- Levis, J. (2007). Computer technology in teaching and researching pronunciation. *Annual Review of Applied Linguistics*, 27, 184–202. <https://doi.org/10.1017/S0267190508070098>
- Levis, J., & Pickering, L. (2004). Teaching intonation in discourse using speech visualization technology. *System*, 32(4), 505–524. <https://doi.org/10.1016/j.system.2004.09.009>
- Lewis Johnson, W. (2010). Serious use of a serious game for language learning. *International Journal of Artificial Intelligence in Education*, 20(2), 175–195. <https://doi.org/10.3233/JAI-2010-0006>
- Lin, C.-H., Warschauer, M., & Blake, R. (2016). Language learning through social networks: Perceptions and reality. *Language Learning & Technology*, 20(1), 124–147. Retrieved February 22, 2018, from <http://llt.msu.edu/issues/february2016/linwarschauerblake.pdf>
- Lord, G. (2015). I don't know how to use words in Spanish: Rosetta Stone and learner proficiency outcomes. *Modern Language Journal*, 99(2), 401–405. https://doi.org/10.1111/modl.12234_3
- Mackey, A., & Choi, J.-Y. (1998). Review of TriplePlayPlus. *Language Learning & Technology*, 2(1), 19–20. Retrieved February 22, 2018, from <http://llt.msu.edu/vol2num1/reviews/tripleplay.html>
- Massaro, D. W., & Light, J. (2004). Using visible speech for training perception and production of speech for hard of hearing individuals. *Journal of Speech, Language, and Hearing Research*, 47(2), 304–320. [https://doi.org/10.1044/1092-4388\(2004/025\)](https://doi.org/10.1044/1092-4388(2004/025))
- Mich, O., Neri, A., & Giuliani, D. (2006). The effectiveness of a computer assisted pronunciation training system for young foreign language learners. *Proceedings of CALL 2006* (pp. 135–143). Antwerp, Belgium: University of Antwerp. Retrieved February 22, 2018, from <http://lands.let.ru.nl/literature/neri.2006.4.pdf>
- Mompean, J. A., & Fouz-González, J. (2016). Twitter based ELF pronunciation instruction. *Language Learning & Technology*, 20(1), 166–190. Retrieved

- April 22, 2018, from <http://lt.msu.edu/issues/february2016/mompean-fouzgonzalez.pdf>
- Movellan, J. R., Eckhardt, M., Virnes, M., & Rodriguez A. (2009). Sociable robot improves toddler vocabulary skills. In *Proceedings of the 4th ACM/IEEE Human Robot Interaction*, La Jolla, CA (pp. 307–308). New York, NY: ACM. <https://doi.org/10.1145/1514095.1514189>
- Nagano, K., & Ozawa, K. (1990). English speech training using voice conversion. In *1st International Conference on Spoken Language Processing (ICSLP 90)*, Kobe, Japan (pp. 1169–1172). Retrieved February 22, 2018, from http://www.isca-speech.org/archive/icslp_1990/i90_1169.html
- Nardi, B., Ly, S., & Harris, J. (2007). Learning conversations in World of Warcraft. In *40th Annual Hawaii International conference on system sciences* (pp. 1–10). Washington, DC: IEEE Computer Society. Retrieved February 22, 2018, from <http://evols.library.manoa.hawaii.edu/bitstream/handle/10524/1601/Nardi-HICSS.pdf?sequence=1>
- Neri, A., Cucchiari, C., & Strik, H. (2006). Selecting segmental errors in L2 Dutch for optimal pronunciation training. *International Review of Applied Linguistics*, 44(4), 357–404. <https://doi.org/10.1515/IRAL.2006.016>
- Neri, A., Cucchiari, C., Strik, H., & Boves, L. (2002). The pedagogy-technology interface in computer-assisted pronunciation training. *Computer-Assisted Language Learning*, 15(5), 441–467. <https://doi.org/10.1076/call.15.5.441.13473>
- Neri, A., Mich, O., Gerosa, M., & Giuliani, D. (2008). The effectiveness of computer assisted pronunciation training for foreign language learning by children. *Computer Assisted Language Learning*, 21(5), 393–408. <https://doi.org/10.1080/09588220802447651>
- Neumeyer, L., Franco, H., Digalakis, V., & Weintraub, M. (2000). Automatic scoring of pronunciation quality. *Speech Communication*, 30(2-3), 83–93. [https://doi.org/10.1016/S0167-6393\(99\)00046-1](https://doi.org/10.1016/S0167-6393(99)00046-1)
- Oh, Y. R., Yoon, J. S., & Kim, H. K. (2007). Acoustic model adaptation based on pronunciation variability analysis for non-native speech recognition. *Speech Communication*, 49(1), 59–70. <https://doi.org/10.1016/j.specom.2006.10.006>
- Ouni, S., Cohen, M. M., & Massaro, D. W. (2005). Training Baldi to be multilingual: A case study for an Arabic Badr. *Speech Communication*, 45(2), 115–137. <https://doi.org/10.1016/j.specom.2004.11.008>
- Park, S., Han, J., Kang, B., & Shin, K. (2011). Teaching assistant robot, ROBOSEM, in English class and practical issues for its diffusion. In *Proceedings of IEEE A Workshop on Advanced Robotics and its Social Impacts*. Retrieved February 22, 2018, from <http://www.arso2011.org/papers>

- Peachey, N. (2009). 20 WebCam activities for EFL ESL students. *Nik's Learning Technology Blog*. Retrieved February 22, 2018, from <https://nikpeachey.blogspot.co.uk/search?q=video+>
- Pearson Education Inc. (2011). Versant English Test. Test description and validation summary. Retrieved February 22, 2018, from <https://www.versant-test.com/technology/VersantEnglishTestValidation.pdf>
- Pearson, P., Pickering, L., & Da Silva, R. (2011). The impact of computer assisted pronunciation training on the improvement of Vietnamese learner production of English syllable margins. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 2nd Pronunciation in Second Language Learning and Teaching Conference*, September 2010 (pp. 169–180). Ames, IA: Iowa State University.
- Pennington, M. C. (1990). The context of L2 phonology. In H. Burmeister & P. L. Rounds (Eds.), *Variability in second language acquisition*, Vol 2 (pp. 541–564). Eugene, OR: University of Oregon. Available at https://books.google.com/books/about/Variability_in_second_language_acquisiti.html?id=wkYqAQAIAAJ
- Pennington, M. C. (1992). Discourse factors in second language phonology: An exploratory study. In J. Leather & A. James (Eds.), *New sounds 92, Proceedings of the 1992 Amsterdam Symposium on the Acquisition of Second-Language Speech*, University of Amsterdam, April 1992 (pp. 137–155). Amsterdam: University of Amsterdam.
- Pennington, M. C. (1999). Computer-aided pronunciation pedagogy: Promise, limitations, directions. *Computer Assisted Language Learning*, 12(5), 427–440. <https://doi.org/10.1076/call.12.5.427.5693>
- Pogue, D. (2004). Speaking naturally, anew. Retrieved February 22, 2018, from <https://www.nytimes.com/2004/12/02/technology/circuits/02POGUEEMAIL.html?ex=1259730000&en=339f1ddcdf5fab69&ei=5088&partner=rssnyt>
- Rama, P., Black, R., van Es, E., & Warschauer, M. (2012). Affordances for second language learning in world of Warcraft. *ReCALL*, 24(3), 322–338. <https://doi.org/10.1017/S0958344012000171>
- Rogerson-Revell, P. (2010). Phonology and phonetics review. Subject Centre for Languages, Linguistics and Area Studies, Higher Education Academy. Retrieved February 22, 2018, from https://www.llas.ac.uk/sites/default/files/mbfiles/mb081/page_01.htm
- Rogerson-Revell, P. (2011). *English phonology and pronunciation teaching*. London: Bloomsbury.
- Scharenborg, O. (2007). Reaching over the gap: A review of efforts to link human and automatic speech recognition research. *Speech Communication*, 49(5), 336–347. <https://doi.org/10.1016/j.specom.2007.01.009>

- Schultz, T., & Kirchhoff, K. (Eds.). (2006). *Multilingual speech processing*. San Diego & London: Academic Press.
- Seedhouse, P. (Ed.). (2017). *Task-based language learning in a real world digital environment: The European digital kitchen*. London: Bloomsbury.
- Sumbly, W., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, 26(2), 212–215. <https://doi.org/10.1121/1.1907309>
- Surface, E., & Dierdorff E. (2007). *Special operations language training software measurement of effectiveness study: Tactical Iraqi study final report*. Special Operations Forces Language Office, Tampa, FL.
- Swain, M. (2000). The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. In J. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 97–114). Oxford: Oxford University Press.
- Swain, M., & Lapkin, S. (1998). Interaction and second language learning: Two adolescent French immersion students working together. *Modern Language Journal*, 82(3), 320–337. <https://doi.org/10.1111/j.1540-4781.1998.tb01209.x>
- Sykes, J. M. (2013). Synthetic immersive environments and second language pragmatic development. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 5549–5505). London: Blackwell.
- Thorne, S., Fischer, I., & Lu, X. (2012). The semiotic ecology and linguistic complexity of an online game world. *ReCALL*, 24(3), 279–301. <https://doi.org/10.1017/S0958344012000158>
- Wachowicz, K. A., & Scott, B. (1999). Software that listens: It's not a question of whether, it's a question of how. *CALICO Journal*, 16(3), 253–276. <http://dx.doi.org/10.1558/cj.v16i3.253-276>
- Wang, X., & Munro, M. J. (2004). Computer-based training for learning English vowel contrasts. *System*, 32(4), 539–552. <https://doi.org/10.1016/j.system.2004.09.011>
- Weizenbaum, J. (1966). ELIZA—A computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), 36–45. <https://doi.org/10.1145/365153.365168>
- Wik, P., & Hjalmarsson, A. (2009). Embodied conversational agents in computer assisted language learning. *Speech Communication*, 51(10), 1024–1037. <https://doi.org/10.1016/j.specom.2009.05.006>
- Wildner, S. (2002). Learn German now! Software review. *CALICO Journal*, 20(1), 161–174. Retrieved January 1, 2018, from https://web.archive.org/web/20021026162807/www.calico.org/CALICO_Review/review/germanow00.htm



6

Assessing Pronunciation

Introduction

As we have stressed in the preceding chapters, pronunciation is a complex area of language proficiency and performance linked to a wide range of human behaviors and incorporating both physiological and social factors. This complexity presents challenges in pronunciation assessment in terms of what and how to measure this aspect of language proficiency. In addition, the fact that pronunciation can vary in relative independence from other aspects of language proficiency raises issues regarding the weight which should be given to an L2 learner's pronunciation in the overall assessment of language competence or achievement. At the present point in time, reconsiderations of the nature of pronunciation and its contribution to language proficiency are having an impact on language assessment, an area of pronunciation theory and application that has developed over time as a tradition constituting a variety of practices connected to language teaching, testing, and research. Pronunciation experts and testing specialists are only just now addressing questions about what is to be assessed and how assessment is to be

carried out that were first raised in the 1990s but have become more pressing as globalization has increased the demand for testing of language competence for academic study and for specific types of employment.

In this chapter, we review testing concepts and the current state of pronunciation assessment and critique a number of standardized tests that include a pronunciation component while also seeking to provide our own perspectives on the “what” and “how” questions of pronunciation testing. We conclude with recommendations for research and practice that incorporate others’ critiques of and recommendations for pronunciation testing along with our own. In particular, we suggest further developing pronunciation assessment to include, on the one hand, more narrowly or autonomously oriented tasks involving auditory perception and imitation and, on the other hand, more broadly communicative tasks involving adjustment of pronunciation in response to factors of audience and context.

The Nature of Pronunciation Assessment

We start from the position that for purposes of assessment, pronunciation cannot logically be defined only or mainly in terms of what a person knows, a conceptual domain, but must be defined in terms of what a person does, as a functional domain of skill, ability, or competence that can be observed in performance. Whether defined in relation to the critical period or social interaction, pronunciation is clearly not only cognition, that is, an organized body of knowledge stored in the brain that can be deliberately accessed. While there is clearly a cognitive dimension to pronunciation, as there is in any form of behavior, in terms of the stored information and network of connections required for acting, as we discussed in Chap. 2, the competence underlying a speaker’s pronunciation cannot be adequately displayed and assessed as a body of knowledge but can be effectively displayed and assessed only through the behavior itself. In language learning, pronunciation is considered to be a domain of skill or ability that is an aspect of language proficiency. Pronunciation is often a facet of speaking assessment geared

to the assessment of non-native or L2 speech, which is the focus of this chapter. Pronunciation assessment is also an aspect of diagnosing phonological impairment in children and adults, as reviewed in Chap. 7. In the practice of so-called “accent reduction” for speakers seeking to shift their pronunciation from a non-prestige to a prestige accent, some form of assessment is likely, though, as far as we are aware, there are no pronunciation tests geared specifically for this purpose.¹ Pronunciation may also be assessed in and of itself, as an autonomous aspect of language proficiency with its own facets, such as assessing pronunciation aptitude in production by imitation tasks, or pronunciation aptitude in perception by identification or labeling of minimal pair distinctions or by recognition of sound-symbol correspondences (see Chap. 2 and below).

Assessing a complex, multi-dimensional human ability or skill set such as spoken language competence and pronunciation specifically is a matter of sampling the domain of interest. According to current practices, the assessment of pronunciation may be focused on intelligibility, fluency, accuracy, nativelikeness, or communicative effectiveness, and the domain of interest may include the assessment of a speaker’s ability to convey different pragmatic and discourse functions through prosody. Pronunciation assessment, including in classroom settings, also sometimes includes perceptual measures such as the ability to discriminate between different phonemes or stress patterns, or to segment speech into information units based on prosody rather than on grammar. Two different ways of assessing and scoring pronunciation are currently in practice, judgements of pronunciation by human raters and assessment of some facets of pronunciation through computerized procedures. In either case, the question of the standard or model against which pronunciation is judged is a crucial issue that relates to the “what” of assessment—its definition, and so its conceptual/theoretical basis and rationale—as well as to the “how” of assessment—the tasks used to elicit evidence of a person’s ability and the procedures for assessing that ability.

The assessment of pronunciation as a productive aspect of language proficiency is a form of **performance-based assessment** in which L2 learners are prompted to speak and then their pronunciation is evaluated based on the speaking task or tasks they have performed. As Lai (2011) noted, “performance-based tasks provide more direct measures of student abilities

than multiple-choice items” (p. 1), though perceptual tests based on multiple choice or other procedures involving selecting the correct answer, such as forced choice, are possible as an aspect of pronunciation assessment (see below). In addition, in a multiple-choice type assessment, Messick (1996) observes, “the selected option can only be appraised for correctness or goodness with respect to a single criterion. There is no record, as in the typical performance assessment, of an extended process or product that can be scored for multiple aspects of quality” (p. 2). Testing realized through discrete-item and limited-choice response is the usual approach for measuring a conceptual domain, where it is easy to quantify the result of measurement on a scale of 0–100% correct response. Performance assessment is less easily quantified and may involve more qualitative and holistic forms of evaluation. When quantification is employed in performance assessment, it is usually based on a scale with a relatively narrow range, typically 1–10 or 1–6, with possible gradations within each level of the rating scale.

Shavelson, Baxter, and Gao (1993) describe performance assessment as involving a number of factors or facets: “we view a performance assessment as a sample of student performance drawn from a complex universe defined by a combination of all possible tasks, occasions, raters, and measurement methods” (p. 216). Because differences in all of these factors can result in differences in the outcome of assessment, the contribution of each one must be carefully considered in the design of assessment tasks and carefully managed in assessment procedure. Issues in the design of pronunciation assessment tasks, whether in an independent evaluation or as a component of a broader assessment of speaking, include those identified by Fulcher (2015): “rating scale development, construct definition, operationalization, and validation” (p. 199). All of the test design issues identified by Fulcher include considerations of the types of tasks used to sample performance, such as how natural and representative they are, and the way the resulting performance is assessed, such as the levels and criteria of rating scales, their appropriateness and consistency, and how easy or difficult they are for raters to use. Further issues that arise in assessment procedure concern the uniformity or consistency of implementation, including the conditions under which the assessment tasks are carried out and the characteristics and performance of the raters and any instruments involved.

Reliability and Validity of Assessment in Standardized Tests

Reliability is the essential foundation of any test worthy of the name and so represents a basic condition of test validity, which is a basic requirement of any well-constructed test (Brown, 2005). Those who design standardized tests strive for high reliability, such as by testing small items of information with little or no variation in possible response—for example, test items that can only be correct or incorrect (e.g., true/false items), or that have one correct answer out of a limited set of choices (e.g., multiple-choice items). In the area of pronunciation, such discrete-item forms of assessment have traditionally been applied to L2 speakers' ability to produce minimal pair distinctions in listen-and-repeat or reading tasks focused on word lists or on sentences or paragraphs specially constructed to contain minimal pair words, judged by a teacher or other rater as correct/incorrect or as one phoneme or another. They have also been applied to speakers' perceptual discrimination or identification of phonemes spoken individually or in minimal pair words as either correct or incorrect or as a choice of one phoneme or another.

In the view of cognitive phonology (Couper, 2011, 2015; Mompean, 2006, 2014), discrimination and identification abilities, such as the perception of phonological categories and boundary cues in phrasing and connected speech, are an important aspect of pronunciation learning and ability underlying speech performance. The testing of these aspects of perception, which lend themselves to discrete-item measurement, can therefore usefully be included as an aspect of pronunciation assessment—though they should not be the only or the primary form of assessment since the ability to perceive distinctions in pronunciation does not automatically predict the ability to produce those same distinctions. Although common in classroom testing, perceptual aspects of pronunciation, as far as we are aware, are not part of any standardized language proficiency test, though these aspects are assessed in some of the scales or subtests of language aptitude tests (see Chap. 2 and below).

As reviewed in Chap. 5, computers have made it possible to precisely and automatically analyze and measure features of sound waves, thus opening up the possibility of using them in the assessment of pronunciation in

individual words or longer stretches of speech. The advantage of using computerized speech analysis tools is the high precision and consistency of measurement they provide. Although machine measurements are close to 100% consistent, as long as the machine is working properly, there can be problems of reliability, based on the limited accuracy of automatic speech recognition (ASR) in terms of error detection, matching speech input to a model, analyzing prosody, and handling certain types of tasks (Chap. 5). In addition, as for any type of discrete measurement, issues of validity arise. Like all other uses of computers, the assessment of pronunciation using computerized speech analysis is only as good as the programming which guides the machine to make its measurements. The computer cannot determine what should be measured within the complex performance domain that is pronunciation; that is the job of pronunciation specialists working with language testers and computer programmers. Thus, while computer-based assessment can be highly reliable at the level of the discrete measurements which the machine records and tallies, the selection of valid measures of pronunciation ability or skill and the equating of machine measurements to pronunciation scores or levels of proficiency are complex matters. Pronunciation as an aspect of speaking performance is commonly assessed in standardized tests by human raters or by a combination of human and machine ratings.

Although spoken language has been extensively assessed since the 1950s (see below), there has been hardly any attention to validity issues specifically focused on pronunciation (rather than speaking more generally), and as a consequence, there are many problems with existing pronunciation measures within standardized tests (Harding, 2017; Levis, 2006; Trofimovich & Isaacs, 2017). Winke (2011), following Messick (1996), argues for a broad conception of validity for language tests so that test validation incorporates not only reliability, concurrent validity, and predictive validity, but also “consequential validity” (p. 633), meaning an assessment of test consequences, specifically, whether a test “has positive impacts” (p. 632). Messick (1996) took a wide view of validity as an ongoing process of collecting and evaluating empirical and theoretical evidence to “support the adequacy and appropriateness of interpretations and actions based on test scores or other modes of assessment” (p. 1).

In Messick’s (1996) conception:

Validity is not a property of the test or assessment as such, but rather of the meaning of the test scores. These scores are a function not only of the items or stimulus conditions, but also of the persons responding as well as the context of the assessment. In particular, what needs to be valid is the meaning or interpretation of the scores as well as any implications for action that this meaning entails.... [V]alidity is an evolving property and validation a continuing process. (p. 1)

From this perspective, test validation is evidence-based and also requires an “argument-based” process of building a case for the value of the test in terms of its uses and consequences, as first proposed by Kane (1990). The concept of the ongoing, evidence-based and argument-based nature of test validation applies to the testing of pronunciation, which is being increasingly analyzed and problematized, even while traditional componential forms of validation are still being applied and accepted in the validation of standardized tests of spoken language proficiency.

Considerations in Assessing Pronunciation

The assessment of pronunciation in speech production incorporates a recognition of the constraints arising from the processing of speech in real time (Bygate, 1987), which means that a more proficient L2 speaker will have more automatized speech production capabilities and will be more able to manage cognitive load than a less proficient speaker, and so more able to speak without disruptions of communication. Less proficient speakers are expected to have less fluent and more errorful speech performance, including more hesitations, false starts, and mistakes (de Bot, 1992; Kormos, 2006; Poullisse, 1999), as well as more likelihood of miscommunication and communication breakdown (Chap. 1). The assessment of pronunciation incorporates a recognition as well that an L2 speaker will have less difficulty with pronunciation when task difficulty or task complexity is lower and greater difficulty with pronunciation when task difficulty is higher. Thus, more difficult or complex tasks are expected to differentiate the pronunciation of speakers at different levels of proficiency.

Skehan (2001) notes the “trade-off” in performing language tasks between limited attentional capacity and different kinds of task demands. According to Skehan (2007, 2014), integrating, linking, or transforming information raises task complexity, as does pre-planning. Skehan (2007, 2014) further observes that structured tasks and knowledge of post-task activities lead to greater accuracy, and familiarity and personal relevance increase accuracy and fluency of tasks, as does planning, with a much greater and more consistent effect of planning on fluency than on accuracy. Skehan (2014) also notes that time pressure lowers performance and that interactivity generally improves performance as compared to monologic tasks. All of these factors are relevant to the design of tasks for pronunciation assessment.

For pronunciation, task difficulty or complexity would seem to be lowered if the task focuses on mechanical production or reproduction (i.e., imitation) such as reading or repeating words or short sentences aloud, or on minimal pair discrimination or identification through multiple or dichotomous choice, since these are highly structured tasks that do not involve any kind of manipulation of information or pressure to communicate. It may seem obvious that pronunciation should be assessed by these kinds of mechanical tasks, which focus the test taker’s attention on language form and so on pronunciation (i.e., on clarity, intelligibility, and accuracy) rather than more complex aspects of meaningful communication. There is also a valid measurement reason to do so, as restricting the task which a test taker is asked to perform makes the performance more uniform and predictable. Limiting and focusing the test taker’s response helps to limit and focus measurement and so to ensure high reliability.

However, there is an issue with the validity of limited, pronunciation-focused, mechanical or reproductive tasks, such as reading or repeating words or short sentences, since this is rarely part of real communication involving speaking, and a person’s pronunciation in a limited, form-focused context will normally differ significantly from pronunciation in running speech. In addition, asking test takers to read words or sentences aloud ties pronunciation to orthography and written language, which may elicit a type of performance that is self-conscious, formal, or unnatural in comparison to the speaker’s pronunciation under other kinds of

conditions. Moreover, a speaker's pronunciation of words and sentences has to do with the speaker's familiarity with the test words and linguistic structures used in test sentences, so that the assessment of pronunciation implicates other aspects of language proficiency, including the test taker's level of literacy, which has been shown to be an important factor determining speakers' ability in linguistic imitation and focus-on-form tasks (Tarone, Bigelow, & Hansen, 2009). The fact that pronunciation is manifested in words and larger units of speech and that it is one among the competing demands for a speaker's internally and externally directed attention and effort suggests that high validity can only be assured by weighing pronunciation among other measures of spoken language performance. Yet there is sometimes an interest in assessing pronunciation, such as intelligibility, accuracy, or phonological aspects of fluency, outside the context of other aspects of speaking proficiency.

The more open-ended and complex the speaking task that the test taker is asked to perform, the more it might be able to differentiate between speakers of different proficiency levels. On the other hand, the more difficult the task, the less predictable and uniform the speech that will be produced and so the more complex the assessment. In considering how to assess pronunciation, there is thus a trade-off between the complexity, and so the validity, of the task used to measure pronunciation and the ease, and so the reliability, of measurement. Validity is increased as the elicited performance moves closer to the type of communication that might occur outside a testing context, and hence to a sample of "actual L2 proficiency" rather than "language-like behavior" (Abrahamsson & Hyltenstam, 2009, p. 254). Validity is therefore increased if assessment of pronunciation incorporates interaction and the ability of speakers to adjust pronunciation if they think the listener has not understood, as was illustrated in Labov's (1966) New York City department store "fourth floor" experiment referred to in Chap. 1. The assessment of speaking skill, and hence of pronunciation, should arguably include what Young (2000) described as *interactional competence*, which includes the factors not only of context and situation, but also of more than one participant in communication as affecting performance. Incorporating an audience factor, potentially including other test takers, test administrators, or additional people as interactants, and the possibility for some

speaker control of the form and content of talk, makes for a more valid and also a more challenging spoken language test, both from the point of view of the test designer and the test taker.

Since the 1990s, applied linguists (e.g., Riggenbach, 1998; Tarone, 1998) have highlighted context-dependent variation affecting both L1 and L2 speech and have called for spoken language tests to include a variety of types of speaking tasks in order to sample a range of speaker competencies and so to increase the validity of those tests. As Chalhoub-Deville (2003) has argued, the context of communication activates different aspects of a language user's abilities and repertoire, so that performance is affected by aspects of the context in which that performance is elicited. This fact has been noted in pronunciation assessment, where there is increasing awareness of the different types of performance that are elicited by different types of tasks, such as reading words, sentences, or a paragraph aloud; responding to specific prompts; communicating in a meaningful task; or spontaneous speech outside a testing context. Shavelson et al. (1993) noted that inconsistencies in performance across tasks, known as *task-sampling variability*, has been a problem in studies of writing, mathematics, and science achievement, where “[l]arge numbers of tasks are needed to get a generalizable measure” (p. 217). The same is true for speaking performance and for the pronunciation component of speaking. Ideally, assessment tasks should assess learners' pronunciation elicited under different conditions, to include some range of variation. This would mesh well with an understanding of pronunciation as being constrained by the requirements of both cognitive processing and interactive communication (what Bygate, 1987, calls “reciprocity”). It also meshes well with sociolinguistic and translanguaging views of pronunciation (Pennington, 2015) as partly automatized and partly under speakers' control, an aspect of their expression of identity, affiliation, and other aspects of meaning.

Measures of Pronunciation

Since the definition of pronunciation is not a settled matter, issues of the construct of pronunciation arise in considerations of assessment. Even as the matter of what pronunciation is remains undetermined, language

teaching, research, and testing have evolved various orientations to the assessment of pronunciation. Two orientations common in all three spheres (teaching, research, and testing) focus on the speaker: accuracy or nativelikeness, which identifies a target for correct performance, and fluency, which identifies skill as automaticity of performance. Two other orientations common in research and testing focus on the listener: intelligibility, which assesses the extent to which a listener understands what a speaker is saying (e.g., through recognizing specific words and other linguistic elements in an utterance), and comprehensibility, which assesses how difficult (or easy) a listener finds it to understand what the speaker is saying. These different orientations to pronunciation, as reviewed in Chap. 1, are the subject of much current debate and discussion regarding the nature and the assessment of pronunciation.

Accuracy, Accentedness, or Nativelikeness

Pronunciation can be seen as a characteristic of performance according to the norms of a certain language or speech community. From this perspective, it is assessed by judgements or goodness of fit measures of the match or the distance between a speaker's pronunciation and that of a model for the language or speech community considered to be the baseline. In this approach, pronunciation might be assessed by evaluating whether a speaker's consonant articulations, vowel articulations, stress, and intonation are correct or accurate; how accurate or nativelike these individual facets of pronunciation are; or how nativelike the speaker's accent is overall. A criterion of correctness or accuracy implicitly assesses pronunciation according to a model language group or speech community whereas a criterion of nativelikeness or accentedness does so explicitly. Error-free, accent-free,² or nativelike performance has long been considered the ultimate goal of language learning. Since small degrees of divergence from a pronunciation target may be "covert contrasts" that are below the level of perception, machine ratings of pronunciation components in relation to a model have higher reliability than human ratings of individual facets of pronunciation. They also have higher reliability given the difficulty human raters have in differentially assessing facets of pronunciation

(vowels, consonants, and features of prosody) and even in differentiating assessments of pronunciation from assessments of other aspects of spoken proficiency such as vocabulary and grammar. However, from the point of view of what the assessment is measuring, a holistic human rating of nativelikeness can be seen to have high validity, as this is an actual judgement people might make about how a speaker sounds. Research by Saito, Trofimovich, and Isaacs (2016, 2017) demonstrates that native speaker judgements of linguistic nativelikeness, equated by them to *accentedness*, are valid measures of pronunciation, as they are strongly linked for adult L2 speakers to segmental phonology, and prosodic variables figure also in those judgements at all levels of proficiency.

For both machine and human ratings, the model underlying the ratings is an issue. There is a great difficulty in defining the baseline in terms of native speaker performance because, as Davies (2017) observes, there is no one standard that can be agreed for native speaker performance: “*the native speaker does not exist*” (p. 185, emphasis in original). Clearly, the standard will differ according to which speech community (e.g., British, American, etc.) the model is based on. As Davies notes, “in the absence of an adequate description of the native speaker, what takes its place is the Standard Language” (p. 186), which is problematic for pronunciation assessment since the latter is based on the written language (p. 187). Often, pronunciation instruction, as discussed in Chap. 3, and assessment aim to emulate an educated speaker’s pronunciation, a “prestige accent,” within a specific speech community, such as “educated Australian English.” In pronunciation assessment, such a standard is usually assumed but rarely overtly described, raising issues of both reliability and validity. There is also, as Davies (2017, p. 186) points out, a political issue having to do with attributions of power and status to speakers according to how their speech is judged in relation to a prestige accent. This raises issues as well of ethics and the consequences of language testing, thus the important concern of consequential validity.

Issues of status and power attendant on judging pronunciation according to a native speaker prestige accent tied to a specific speech community can be addressed by defining the standard on English as an international language or lingua franca, as Jenkins (2000) has advocated. The whole enterprise of evaluating a person’s pronunciation according to

native speaker norms can be questioned as a way to assess proficiency in spoken language since a person can achieve high functionality in a language while maintaining a degree of distance from native speaker norms in pronunciation. Therefore, some argue for moving away from this approach to pronunciation assessment and instead focusing on different kinds of measures, such as fluency, intelligibility, or comprehensibility. In addition, we have proposed in Chap. 1 a conceptualization of pronunciation competence, which may be referenced to English as an international language or lingua franca, as Ur (2012), following Jenkins (2000), suggests; or it may be referenced to specific contexts of communication, such as specific types of employment and job requirements involving pronunciation skill. However, the issue remains of who is to be the judge of these other measures of pronunciation.

Fluency

Fluency is considered to be a global type of proficiency indicative of skilled communicative performance as a result of learning and automatization of behaviors over time, as observed in the speaker's ability to produce continuous and coherent speech in phrasal, clausal, and discoursal units, and to avoid disfluencies and communication breakdowns (Chap. 1). Segalowitz (2010) differentiates a speaker's **utterance fluency**, those fluency features of speech production which can be measured instrumentally, from the underlying competency assumed to be responsible for utterance fluency, which he labels **cognitive fluency**, "the efficiency of operation of the underlying processes responsible for the production of utterances" (p. 165). In Segalowitz's (2010) view, listeners' judgements of a speaker's fluency, that is, **perceived fluency**, are "the inferences listeners make about speakers' cognitive fluency based on their perceptions of their utterance fluency" (p. 165). From this perspective, cognitive fluency would be the underlying construct that is assessed by proxy, through either mechanical means or listener judgements, or a combination of the two.

Utterance fluency in Segalowitz's sense, or what Lennon (1990) and others have referred to as "temporal fluency," involves measures of the timing, or speed, and continuity of speech, such as speaking rate and the proportion of

pausing and non-lexical filler. A speaker's fluency in speaking performance might be assessed by measuring the rate and quantity of speech produced in a given period of time, such as the number of syllables or words produced, or the average length of unbroken stretches of speech. It might also be assessed by counting the number and length of silent pauses and other pauses filled with hesitators such as *er* or *uh*, or by taking an average of these within a stretch of speech. In this narrow way of conceptualizing it, fluency would be evaluated and measured by considering the number and size of continuous and discontinuous units (i.e., pauses) produced in speech.

The timing and continuity of speech affect not only fluency in general but pronunciation in particular, as captured in the notion of "phonological fluency" (Pennington, 1989, 1990, 1992), which incorporates coarticulation, weakening of articulation and stress, and grouping of elements by prosodic means under conditions of sustained production of speech. As far as we are aware, there have been no attempts to focus fluency measures directly on pronunciation features such as those identified by Pennington (1989, 1990, 1992). However, temporal fluency is often used as an indirect measure of pronunciation that relates to more global measures of speaking proficiency. Bosker, Pinget, Quené, Sanders, and de Jong (2012) examined different types of fluency measures for L2 Dutch which could be applied by machine or human raters, categorizing them into three types: "*breakdown fluency* concerns the extent to which a continuous speech signal is interrupted; *speed fluency* has been characterized as the rate and density of speech delivery; and *repair fluency* relates to the number of corrections and repetitions present in speech" (p. 160, emphasis in original). They found that acoustic measures of pausing or breakdown and of speed of speech best predicted untrained raters' perceptions of L2 speakers' fluency, with repair measures contributing little to judgements of fluency. Bosker et al.'s (2012) result can thus help to narrow down the features of speech that should be incorporated into automated assessments of fluency.

Intelligibility

Intelligibility, defined as the ability to be understood (Munro, Derwing, & Morton, 2006, p. 112) or to speak in a way that linguistic components can be recognized by a listener (Smith & Nelson, 1985, p. 334), is a criterion

that disregards accent, as long as a person's words and longer stretches of speech are understandable. It is thus a relevant measure of pronunciation for both L1 and L2 audiences, as a level of pronunciation clarity that is required for communication to take place and to avoid miscommunication based on failure to produce phonemes and prosody in discriminable ways. Intelligibility can be measured by having listeners perform an orthographic transcription of a speaker's utterance, to see if listeners agree on what was said, or to check if the transcript matches what the speaker was trying to say (e.g., based on a read script). Alternatively, intelligibility may be tested by having raters fill in gaps in a speaker's transcript, as done in Browne and Fulcher's (2017) research.

Intelligibility can be seen as establishing a crucial pronunciation baseline that is a necessary condition for communication. The question is whether it is a sufficient condition that can stand as a criterion measure for pronunciation of English as an international language (Cruttenden, 2014) or lingua franca (Jenkins, 2000), or in some circumstances, such as academic study in an English-speaking institution, and types of interactions, such as customer-service encounters or other business and social negotiation.

Comprehensibility

Comprehensibility, the ease/difficulty of understanding of speech (Munro & Derwing, 1995; Munro et al., 2006), is based in part on intelligibility. It incorporates pronunciation while also incorporating lexical and grammatical competencies (Chap. 1). However, intelligibility is sometimes defined in a way that conflates it with comprehensibility. In this connection, Levis (2006) noted what he termed "narrow" and "broad" definitions of intelligibility. The narrow definition is what is generally meant by "intelligibility" as this term has been used by researchers to refer to the ability to analyze an L2 speakers' stream of speech into words and larger utterances. The broad sense in which "intelligibility" is used in relation to pronunciation may include word and utterance recognition but is more essentially about understanding the speaker's meaning, thus "comprehension" or "comprehensibility" in the usual sense of

this word (see Chaps. 3 and 8 for discussion of research using these concepts). This broad notion of intelligibility, which is in essence comprehensibility as measured by listeners' subjective scalar ratings, is the notion underlying many proficiency test scales for spoken language performance, as noted by Isaacs & Trofimovich (2012, p. 477), such as in the Test of English as a Foreign Language (TOEFL) and the International English Language Testing System (IELTS).

Whereas actual comprehension of speech—that is, “intelligibility” in the sense of Munro et al. (2006), as “the extent to which a speaker’s utterance is actually understood” (p. 112)—might be assessed by comprehension questions or true-false statements related to the speaker’s intended message, comprehensibility is generally conceived of as listeners’ perceptions of how easy or hard it is to understand an L2 speaker, evaluated on an ease-of-understanding scale, the approach taken by Munro and Derwing (1995, 2001).

Perception

It is common in classroom assessment to assess perception of pronunciation according to English L1 distinctions, such as the ability to discriminate or recognize specific segmental phonemes (e.g., in minimal pairs) or stress and intonation patterns (e.g., of contrast or emphasis, given or new information). Listening tests may be focused on what might be called “perceptual intelligibility,” through dictation or cloze tests which require perception of morphemes that occur with weak stress and reduced articulation, such as function words (e.g., pronouns; articles *a* and *the*; auxiliaries *is*, *are*, *has*, *have*; and one-syllable prepositions such as *of*, *at*, *in*, and *on*), and suffixes (e.g., *-s*, *-ed*, and *-ing*). Listening comprehension tests indirectly incorporate perceptual ability in phonology when they include words that are potentially confusable or difficult to recognize phonologically or information that is low in intelligibility, such as because of the informality or rapidity of speech.

Thus, the assessment of pronunciation perception, which is often done directly as an aspect of language teaching, is generally done indirectly, via listening comprehension, in standardized language proficiency tests. It can be

noted that the Versant automated test of spoken proficiency (Bernstein, 2004; Bernstein, Van Moere, & Cheng, 2010; Suzuki, Balogh, & Bernstein, 2006), described below, includes sentence repetition, a task which incorporates an element of pronunciation-focused speech perception that underlies intelligibility—specifically, the ability to analyze the speech stream in terms of segmental sounds and sequential patterns. The perceptual processing of speech in terms of its component sounds, sound sequences, and information units marked by prosody (e.g., by stress, tone, and linking) is a minimum requirement for being able to recall and repeat, whether accurately or not, what has been heard. The ability to discriminate or identify individual sounds and sequential pronunciation patterns and to hold them in mind as a basis for repetition/imitation is also tested in some forms of assessment of pronunciation aptitude (see below and also the discussion of pronunciation aptitude testing in Chap. 2).

Pronunciation perception goes far beyond this, however, to the perception and interpretation of both segmental and prosodic cues to information structure and message pragmatics that communicate metamessages and have interactional effects in context. Although these aspects of perception are important dimensions of pronunciation competence, they have to date played almost no role in assessment.

Multiple Measures

It is common in speaking scales to use multiple measures of pronunciation, sometimes combined and sometimes separately rated, such as phoneme accuracy, fluency, and intelligibility. As noted by Harding (2017), the criterion of “Delivery” used in some speaking tests often incorporates a range of pronunciation and fluency features, such as the TOEFL Speaking scale, which includes intelligibility of articulation, intonation, flow, and pace; and other speaking tests use a combination of pronunciation measures in their scales. When more than one measure is used in pronunciation assessment, the question arises as to whether the different aspects of the assessment are independent or not.

Many studies have found that ratings of fluency are not independent of other pronunciation measures, but correlate with them, including experts’ ratings of the quality of speakers’ segmental phonemes, which

have a positive correlation to their ratings of fluency (Cucchiariini, Strik, Binnenpoorte, & Boves, 2002), and of accentedness, which has a negative correlation to fluency (Anderson-Hsieh & Koehler, 1988; Munro & Derwing, 1998, 2001; Derwing, Rossiter, Munro, & Thomson, 2004; Wennerstrom, 2000). In the case of accentedness and fluency, however, the correlation may not be a strong one, and the two constructs may be separable for purposes of pronunciation assessment (e.g., Pinget, Bosker, Quené, & de Jong, 2014). As Pinget et al. (2014) suggest on the basis of their research, “Ratings on ‘delivery’, covering both fluency and accent, may thus prove problematic for raters since the two concepts can very well exist independently from the other” (p. 363).

An important consideration in traditional componential validation of pronunciation measures is the extent to which they correlate with each other and with other measures of language proficiency, such as overall speaking proficiency. High correlations suggest that measures are assessing the same phenomenon. High correlations thus challenge the independence of ratings which are supposed to be independent (e.g., of segmental accuracy and fluency rated separately on a given test) but are a valued measure of validity when ratings according to a new test are compared to those on an accepted test. With the goal of automated language testing, an issue is the extent to which objective or quantitative, machine-ratable measures of L2 speech can be related to raters’ qualitative (scaled or holistic) assessments of the quality or level of L2 speech, such as how highly acoustic measurements of phoneme accuracy or utterance fluency are correlated to subjective judgements of phoneme accuracy, accent, intelligibility, or fluency (i.e., perceived fluency). In research carried out on measures of pronunciation in Dutch, Cucchiariini et al. (2002) found that for both read and spontaneous speech, experts’ ratings of fluency were related to their ratings of the quality of speakers’ segmental phonemes and their overall pronunciation quality ratings as well as to machine measures of temporal fluency, especially rate of articulation of sounds and pause frequency, as noted in their previous work (Cucchiariini, Strik, & Boves, 2000).

However, in the Cucchiariini et al. (2002) study, read speech was more consistently and effectively rated on all scales than spontaneous speech, suggesting that it is easier to assess pronunciation in read

speech than spontaneous speech, and also that measures which are appropriate (reliable and valid) for judging pronunciation in read speech might be different from those which are appropriate for assessing pronunciation in spontaneous speech. A later study within the Dutch research group (van Doremalen, Cucchiarini, & Strik, 2010) revealed a higher frequency of phonetic substitution and deletion errors in read than the spontaneous speech, which may be due to interference from routinized pronunciation patterns tied to written language, or to the stress or unnaturalness which many feel when reading aloud in an academic or testing environment. As van Doremalen et al. (2010) comment, “If some of the errors observed in L2 read speech are simply decoding errors caused by insufficient knowledge of L2 orthography or interference from it, it is legitimate to ask whether such errors are pronunciation errors at all.” Asking this question, in the context of the results of this set of related studies, suggests the complexity of pronunciation assessment, how many issues are just becoming apparent, and how many questions remain to be asked and answered. Questions which have not sufficiently been addressed in relation to multiple measures in pronunciation assessment are (i) which combined features of pronunciation competence are most relevant for specific types of work and ESP contexts and (ii) how these should be assessed.

Refining Measures of Pronunciation in Proficiency Test Development

Test developers continually refine their scales and descriptors, including those for pronunciation, to increase their clarity and usability in making discriminations across scale points and levels and in response to research evaluating their effectiveness (reliability and validity). In the last ten years, researchers have started to focus on how raters assess different elements of pronunciation and whether certain ones are weighted more heavily than others, as Bosker et al. (2012) found for pausing and speech rate as against repair moves in the assessment of fluency. Automated speaking assessment systems are continually being improved in order to

increase the correlations between their ratings and those of human raters, an important consideration for both their criterion-related and face validity that is also central to building an argument about whether the automated assessment matches the reality of people's judgements of L2 speaking competence. If their ratings of pronunciation do not match those of other accepted ways of assessing it, then their consequential validity is at issue in the important sense that the testing of pronunciation has become big business with significant consequences for people's lives and employment.

Assessment of Pronunciation by Human Raters

Working with Human Raters

Assessment of pronunciation by humans may be done by naïve raters who have no relevant specialized knowledge or experience, other than their native speaker status, or by expert raters who have relevant specialized knowledge and experience. The logic behind using naïve native speaker raters is that they would be expected to assess pronunciation in the same intuitive way any native speaker would. The logic behind using expert raters is that they would have skills relevant to pronunciation assessment and so be more likely to assess in a detailed way, according to multiple criteria. Expert raters may be selected for expertise in linguistic phonetics and phonology and/or for their knowledge and experience of non-native pronunciation or the specific second language(s) of test takers. They are sometimes selected specifically for their experience in rating pronunciation by a specific measurement instrument or type of rating scale or task.

Whether expert or non-expert, the assessors must have some test-specific training in order to ensure that they perform their assessments in the manner intended by the test designers and also with consistency from one test taker to the next and from one rater to the next. This initial rater training, which in effect biases the raters to the criteria or categories and measurement tasks of the test, is intended to ensure high reliability of individual raters' scores, that is, high **intra-rater reliability**, and also high reliability across raters, that is, high

inter-rater reliability. An assessment task or test with low reliability of either type cannot be considered to validly assess the domain it is intended to assess.

Fulcher (2015) identifies the reality as well as the difficulties surrounding individual differences in rater response in the assessment of speaking:

Left to their own devices, raters tend to vary in how they score the same performance. The variability decreases if they are trained; and it decreases over time through the process of social moderation. With repeated practice raters start to interpret performances in the same way as their peers. But when severed from the collective for a period of time, judges begin to reassert their own individuality, and disagreement rises. (p. 201)

This suggests the need for raters to work together at one time and in the same place, which limits the possibilities for administering an assessment task or test which is scored by human raters: either the processes of assessment and scoring both take place in the same time and space, or speakers' performance must be audio-recorded so that the scoring can be done at a distance by a group of raters all working together at the same time in another place. A third alternative is to develop a scoring procedure that removes the element of individual human judgement, by audio-recording speakers' performance and analyzing that performance by machine. There is still an element of human judgement, as the interpretation of those analyses must be tied to some meaningful evaluations of performance in terms of useful descriptors or levels of performance or proficiency.

Where there is significant rater variation, even after training, this tends to invalidate a test or assessment procedure, though for tests of complex skills such as writing or speaking, a degree of interrater variation (up to about 30%) is generally expected and accepted. If the interrater variation is systematic, that is, if it is attributable to specific characteristics of raters (other than specific expertise), this calls a rating into question. Besides the complexity of the domain being measured, low inter-rater reliability can be caused by what is termed "construct-irrelevant variance" that is related to listener background variables (e.g., accent familiarity).

Reliability and Validity Issues Involving Rating Scales and Human Raters

As Fulcher (2015) observes of speaking assessment, “Scale development and construct definition are inextricably bound together because it is the rating scale descriptors that define the construct” (p. 199). While there are traditions in terms of the ways that pronunciation is assessed, usually as an aspect of testing spoken language proficiency in a second language, measures of pronunciation have not been designed based on an agreed construct. As a result, the assessment of pronunciation suffers from many problems in test design and implementation that raise issues of reliability and validity. In addition, given that pronunciation is a complex type of behavior that can be measured in different ways, reliability and validity issues can be raised by the rating measures and the specific raters or type of raters employed, as well as by the type of task used to elicit performance and by situational factors affecting performance.

According to Harding (2017), “Pronunciation scales have been shown to be highly problematic to design and implement, with descriptors suffering from inconsistencies, vague language, conflated constructs and unclear trajectories” (p. 12). Inconsistencies often mean that the descriptors for different scale points are not comparable. Inexplicit or vague language means that descriptors do not provide descriptions that can be easily matched to aspects of performance. Descriptors sometimes conflate different aspects of pronunciation, such as accuracy or nativelikeness, intelligibility, and/or fluency, all of which are distinguishable. Unclear trajectories means that the progression from lower to higher scale points is obscure. Scales may have design problems such as being too complex or too simple (too general or brief) to use easily, or they may not incorporate those aspects of pronunciation that are most salient to the raters (Harding, 2017, p. 13). Trofimovich and Isaacs (2017) point to the “growing body of research in language assessment...suggesting that various linguistic measures of L2 pronunciation often fail to distinguish between adjacent levels in multi-level pronunciation scale” (p. 262). Moreover, as Harding (2017) reports, “scales themselves have a limited capacity for ensuring valid interpretation and consistent application among raters.... In recognition of this, ongoing rater training

and other supporting documentation such as benchmarked performances are often recommended to scaffold the role of the scale in the rating process” (p. 13).

Raters’ differential familiarity with speakers’ L1 (Winke, Gass, & Myford, 2012) or with interlanguage phonology can affect their ratings of pronunciation (Carey, Mannell, & Dunn, 2011, p. 204), and, specifically, their ratings of intelligibility and fluency (Derwing et al., 2004). As Browne and Fulcher (2017) have shown, “variation in listener familiarity with L2 speech results in changes to scores on speaking tests...variation [that] is associated with [raters’] estimates of [speaker] intelligibility...” (p. 37). Differences in language background can also affect the features of pronunciation, segmental or suprasegmental, that raters focus on (Kang, Moran, & Vo, 2016). Given that judgements of fluency, intelligibility, segmental sounds, and suprasegmental features depend on raters’ familiarity with specific L1s and types of L2 speech, Browne and Fulcher suggest the value of “accent familiarity training to raters across the range of L1s represented in the test taker population at large” (p. 51).

Inherent sources of variability in rater responses and the administrative difficulties and expense of using human raters have led to increasing interest in machine ratings of pronunciation and speaking proficiency more generally. At the present time, both human and machine ratings are being used, as many of the major testing services and companies offer the option of online testing that digitally records speech samples for later assessment by human raters while also providing a machine-analyzable speech file that can be used in their automated test scoring development.

Assessment of Pronunciation by Machine

The Nature of Automated Assessment of Pronunciation

Since the time of the sound spectrograph when “visible speech” was made possible, new methods of analyzing pronunciation by machine have been developed for research, teaching, and testing. Because of the great advantages in being able to automate the testing process, computers have

increasingly become a preferred system for delivery of tests using the internet. In addition, once test takers' performance data have been captured in digital files, there is the obvious possibility of scoring the tests by computer means. For discrete-item tests, machine scoring is far more efficient and cost-effective than human scoring, and it has the added benefit of avoiding scoring error. Yet even though testing and marking by machine can solve reliability problems associated with human scoring, computerized assessment and scoring of complex forms of performance raises many issues of validity that have not so far been resolved.

The general approach for machine rating of pronunciation is to assess discrete measures of performance (e.g., the acoustic traces of specific words or phonemes), derived from digitized speech samples, against a digitized model of performance according to those same measures. ASR programs, software for acoustic analysis and for timing and counting aspects of the speech signal, and various types of algorithms are used to take the measurements and to make the ratings of test takers according to those measurements. The model against which the assessment of pronunciation is made is established from a database of analyzed samples of native speaker performance, contrasted with a database of analyzed samples of non-native speaker performance to determine the most common characteristics of native versus non-native pronunciation and the degrees of deviation of the latter from the former. The computer analyses of the pronunciation features of the two databases and the differences between them provide data for mathematical algorithms that score performance. The assessment of pronunciation is then a mechanical computation of pronunciation measurements taken by the machine from the test taker's speech sample, comparison of those measurements to the native speaker standards and to the degrees of deviation from those standards found for non-native speakers, and assignment of a score based on those considerations.

Reliability and Validity Issues Involving Machine Assessment of Pronunciation

Reliability issues are rare in machine calculations or tabulations but immediately arise as soon as what the computer is asked to do is more than a simple calculation or tabulation of something in the speech

signal. Questions of reliability can arise if different results are obtained from measurements normed on different databases of native speaker performance (e.g., from different individual speakers or tasks). Validity issues concern the fact that the testing procedure involves no-face-to-face interaction, is 100% mechanical, and samples only limited aspects of performance, with a focus on types of performance that are most readily assessed by machine, such as the quantity of speech produced per unit of time rather than the specific qualitative features of speech.

Browne and Fulcher (2017, p. 38) question the possibility of validity of automated assessment of pronunciation based on matching a set of measurable features of test takers' pronunciation with pre-set norms for performance on constructed tasks such as sentence repetition. One issue is how to set these norms, such as on a sample of native speaker speech. Another is the limited construct of language in terms of meaningful communication when the task is sentence repetition. In the last decade, Scharenborg (2007) and Levis (2007) raised concerns about whether automated pronunciation assessment was necessarily restricted to limited and focused task types that did not represent an adequate range of pronunciation performance, and recently Trofimovich and Isaacs (2017, pp. 265–266) have raised concerns about the ability of automated systems to handle linguistic information other than speaking tasks with “highly predictable” outcomes and “acoustic phenomena that are easy for the machine to score.” It can be noted that there are aspects of pronunciation that can validly be assessed as narrowly specified competencies in auditory perception (e.g., discrimination abilities) and focused production (e.g., repetition or imitation tasks), as discussed below, and automated pronunciation assessment would seem to have some advantages over human ratings of pronunciation in terms of the practicality and fairness concerns of consequential validity. However, more complex aspects of speech are beyond current capabilities of machine rating, though a great deal of effort is being directed at developing and improving ASR and systems for automated speech assessment, towards ambitious goals of assessing speech produced by speakers with the fewest possible controls or constraints (i.e., close to naturally occurring, authentic speech).

Standardized Speaking Tests and Scales

According to Fulcher (2015), oral testing was developed at the Foreign Service Institute (FSI) in the 1950s, precipitated by the Korean War and then expanded into the Defense Language Institute, the Central Intelligence Agency, the Peace Corps, and academia. At present, “The need for speaking tests has expanded from the educational and military domain to decision making for international mobility, entrance to higher education, and employment” (Fulcher, 2015, p. 198). In addition to students’ language proficiency, it is a common practice for language teachers in U.S. public schools to be required to have their English-speaking accent assessed in order to receive teaching certification. In Michigan, for example, as reported by Ballard and Winke (2017, pp. 123–124), certification requires teachers to pass an OPI exam, with the criterion determined by the American Council on the Teaching of Foreign Languages (ACTFL, 2012) Advanced-Low speech descriptor assessing whether “speech can be understood by native speakers unaccustomed to dealing with nonnatives” and “convey their intended message without misrepresentation or confusion” (p. 11). These or other commercial tests of spoken language proficiency may be used for assessment of the oral ability of international teaching assistants (ITAs), and specific standardized tests that include pronunciation measures have been developed for international medical graduates (IMGs).

In most standardized tests, pronunciation is represented as a subsidiary consideration within speaking proficiency, generally defined in terms of multi-factor models of communicative competence. Isaacs (2014) argues that models of communicative competence do not capture the complexity of the contributions of pronunciation to either perception or production of speech, nor to its interaction with other aspects of language such as orthography. In reviewing some of the standardized tests, Levis (2006) refers to a “haphazard collection of descriptors suggest[ing] that pronunciation is relatively unimportant in determining speaking proficiency” (p. 245). In what follows, we briefly review some of the better known and more widely used of these tests.

Some Widely Used Oral Proficiency Tests and Scales

American Council on the Teaching of Foreign Languages (ACTFL) Oral Proficiency Interview and Proficiency Guidelines

The Oral Proficiency Interview (OPI) used by the American Council on the Teaching of Foreign Languages for gathering a spoken language sample is available by telephone with a certified ACTFL tester or by internet with a virtual avatar interviewer, takes 20–40 minutes to complete, and is recorded and assessed by certified raters who score it with reference to either the ACTFL or Interagency Language Roundtable (ILR) Scale. Starting in the 1990s, a number of studies (e.g., Fulcher, 1996; Johnson, 2000; van Lier, 1989; Young & Milanovic, 1992) highlighted the constraints and unnaturalness of the oral proficiency interview context as a factor affecting test takers' performance and the validity of the ACTFL procedures. Since that time, the interview has been modified substantially with the aim of making it more interactive and closer to a natural conversation between two people, though the interview is in fact structured to elicit a speech sample that can be rated according to rating guidelines. Whether the performance obtained by a telephone interview is for purposes of assessment equivalent (or superior) to that obtained in a face-to-face interview, and whether the performance obtained by an online interview with an avatar is equivalent to either of the other modalities is a matter for ongoing validation.

According to the ACTFL website, “the ACTFL Guidelines describe the continuum of proficiency from that of the highly articulate, well-educated language user to a level of little or no functional ability.”³ In the latest (2012) version of the guidelines, the continuum is divided into Novice, Intermediate, and Advanced levels with High–Mid–Low levels and includes as well two higher levels, Superior and Distinguished—thus, 11 identified levels of performance. The Speaking test Guidelines include pronunciation in its language proficiency descriptions but not in any consistent way within or across levels. The extreme variation in the way the Speaking guidelines reference pronunciation—including accent,

accuracy, fluency, and intelligibility—suggests that pronunciation has not been a main consideration in developing the guidelines, which can be criticized as not giving adequate attention to this component of speaking proficiency. It can also be noted that the ACTFL Listening test, as in other standardized language proficiency testing programs reviewed below, is a comprehension test with no specific reference to perceptual pronunciation skills.

Interagency Language Roundtable (ILR) Scale

The Interagency Language Roundtable (ILR) Scale⁴ used in all U. S. government agencies is administered based on an oral interview like the OPI but related to the proficiencies required in specific agencies.⁵ It is administered variably in face-to-face, telephone, video conferencing, and online modes, raising the same issues of test-form equivalence as for the ACTFL OPI. The scale is in the same tradition as the ACTFL scale and has the same number of levels, though differs from it a number of ways. It is unusual in having a zero level of No Proficiency and a top level (Level 5) defined as “Functionally Native Proficiency.” The ILR descriptors provide a considerable amount of detail, but like the ACTFL, the descriptions for pronunciation do not provide clear guidance for rating as they vary significantly across levels and do not show an obvious progression. The descriptions for the bottom three levels refer (with some inconsistency) to specific features of segmental and prosodic phonology, while higher levels refer to pronunciation not at all or to intelligibility, fluency, and/or accuracy. As another type of inconsistency, some descriptions single out pronunciation while others include pronunciation in more complex descriptions of speaking performance.

Common European Framework of Reference (CEFR) Language Proficiency Scale

The Common European Framework of Reference (CEFR) language proficiency scale is a six-level set of language proficiency scales and descriptors produced in the 1990s by the Council of Europe (2001) that is widely used in Europe to provide common reference points for language policy

and practice, including for testing. Thus, the CEFR levels provide reference points for assessing performance on a number of other tests, including IELTS and CAE, described below, as well as other major language tests such as TOEFL. It is encouraging to see such an influential framework affecting the structure of language tests.

The scale for pronunciation, termed the “Phonological Control Scale,” developed in 2001, describes pronunciation proficiency for five of the six CEFR levels, excluding the top level, as shown in Fig. 6.1. The phonological control scale “has been critiqued by researchers as lacking consistency, explicitness and a clear underlying construct” (Harding, 2017, p. 16). Harding (2017) carried out research with experienced pronunciation raters, who found the scales unclear (using terminology that could not easily be matched to performance, such as “natural” to describe intonation and “noticeable” to describe foreign accent), inconsistent across levels (in terms of reference to intonation and

The CEFR Phonological Control Scale
C2 As C1
C1 Can vary intonation and place sentence stress correctly in order to express finer shades of meaning.
B2 Has acquired a clear, natural, pronunciation and intonation.
B1 Pronunciation is clearly intelligible even if a foreign accent is sometimes evident and occasional mispronunciations occur.
A2 Pronunciation is generally clear enough to be understood despite a noticeable foreign accent, but conversational partners will need to ask for repetition from time to time.
A1 Pronunciation of a very limited repertoire of learnt words and phrases can be understood with some effort by native speakers used to dealing with speakers of his/her language group.

Fig. 6.1 The CEFR phonological control scale (Council of Europe, 2001, p. 117)

foreign accent), and with potential irrelevant or tangential descriptors (specifically, mention of “pronunciation of a very limited repertoire of learnt words and phrases”). The conciseness of the scale was generally seen as a plus, but there was some recognition by raters of the trade-off between simplicity and conciseness and sufficient detail for consistent rating decisions. As Harding (2017) comments, “The theoretical divide between pronunciation and aspects of fluency—stress timing, hesitation, ‘chunking’—becomes harder to maintain when human raters, who need to apply scales in practice, struggle to separate these dimensions for judgement purposes” (p. 27). In Harding’s (2017) study, “several raters did not feel comfortable with the view of English that was embodied in the scale, and...some raters at least saw it as out of step with more recent conceptualizations which have problematized the reliance on the native speaker model, and which have highlighted the pluricentricity of English” (p. 28).

The pronunciation descriptors within the CEFR have since been substantially revised, with the recognition that “the phonology scale was the least successful of those calibrated in the original research” (Council of Europe, 2017, p. 132). The new Phonological Control Scale has been substantially expanded and provides a much more nuanced set of descriptors. There are no longer references to “native speaker” in the criteria and the focus has been broadened beyond accent and accuracy to take more consideration of “context, sociolinguistic aspects and learners’ needs” (p. 133). The scale uses three categories with descriptors focused on communicative effectiveness—“overall phonological control,” “sound articulation,” and “prosodic features”—and covering all of the CEFR levels, from A1 to C2. Figure 6.2 shows the descriptors for levels C2 (highest level) and A1 (lowest level).

The development of a more elaborated scale for pronunciation by a highly influential body models the current trend of pronunciation receiving great attention in language teaching and testing and is likely to raise the profile of pronunciation in language teaching and testing not just in Europe but worldwide. It is however too early to know how effective the new phonological scale is in assessing pronunciation competence, but it is encouraging to note the CEFR’s concern “to provide realistic scales and concrete descriptors to support practitioners and

	OVERALL PHONOLOGICAL CONTROL	SOUND ARTICULATION	PROSODIC FEATURES
C2	Can employ the full range of phonological features in the target language with a high level of control – including prosodic features such as word and sentence stress, rhythm and intonation – so that the finer points of his/her message are clear and precise. Intelligibility and effective conveyance of and enhancement of meaning are not affected in any way by features of accent that may be retained from other language(s).	Can articulate virtually all the sounds of the target language with clarity and precision	Can exploit prosodic features (e.g. stress, rhythm and intonation) appropriately and effectively in order to convey finer shades of meaning (e.g. to differentiate and emphasise).
A1	Pronunciation of a very limited repertoire of learnt words and phrases can be understood with some effort by interlocutors used to dealing with speakers of the language group concerned. Can reproduce correctly a limited range of sounds as well as the stress on simple, familiar words and phrases.	Can reproduce sounds in the target language if carefully guided. Can articulate a limited number of sounds, so that speech is only intelligible if the interlocutor provides support (e.g. by repeating correctly and by eliciting repetition of new sounds).	Can use the prosodic features of a limited repertoire of simple words and phrases intelligibly, in spite of a very strong influence on stress, rhythm, and/or intonation from other language(s) he/she speaks; his/her interlocutor needs to be collaborative.

Fig. 6.2 Extract from the CEFR phonological control descriptor scale (Council of Europe, 2017, p. 134)

learners alike, in the delicate and crucial process of acquiring an appropriate and effective pronunciation of the target language” (Piccardo, 2016, p. 6).⁶

International English Language Testing System (IELTS) Speaking Exam

The International English Language Testing System (IELTS)⁷ assesses language proficiency on a scale from 1 (non-user) to 9 (expert). The Speaking exam is conducted face-to-face and includes three parts. In part 1, an examiner asks general questions on a range of familiar topics. In part 2, the test taker is given a card specifying a topic which the test taker is to talk about for 2 minutes, with a minute of preparation time provided and with one or two follow-up questions by the examiner. In part 3, the examiner asks further questions about the topic of part 2 to give the test taker an opportunity to discuss “more abstract ideas and issues.” Performance is rated by examiners according to equally weighted criteria for Fluency and Coherence, Lexical Resource, Grammatical Range and Accuracy, and Pronunciation. According to the public version of its speaking band descriptors, IELTS assessment incorporates pronunciation in a systematic way, with evidence of consistency across levels in reference to “range of pronunciation features,” “flexible” use of these, and “intelligibility.”⁸ However, research has shown that raters find the pronunciation scale the most difficult one to assess (Yates, Zielinski, & Pryor, 2011) and that the descriptors are difficult to apply and do not clearly distinguish different levels of performance for raters (Isaacs, Trofimovich, Yu, & Chereau, 2011). The IELTS pronunciation rating criteria nonetheless stand out for the inclusion of the speaker’s ability to adjust pronunciation and of both quantity and quality criteria for judging pronunciation features, suggesting a richer construct of pronunciation underlying this aspect of assessment in its speaking test as contrasted with most other tests. Also different from other tests, fluency is assessed separately from pronunciation, as a construct linked to coherence, in a conception of fluency that includes temporal and organizational aspects of speech.

Cambridge English: Advanced (CAE) Exam

The Cambridge English: Advanced (CAE) Exam⁹ involves 15 minutes of individual and interactional speaking tasks in four parts, with tasks similar to those for the IELTS Speaking exam (which is also part of the Cambridge suite of English exams), but with two (occasionally three) examinees and two examiners, one of whom takes notes while the other administers the instructions and asks questions. The first part is a brief interview by the examiner with each of the examinees asking them a series of questions designed to elicit information about themselves. In part 2, each examinee selects two pictures to talk about for one minute, followed by the other examinee making a comment on what the previous examinee has said. The third part requires the two examinees to converse about a topic for 2 minutes based on written prompts and then to talk for a further minute in order to come to a decision, based on a further oral prompt by the examiner. The final part is a 5-minute discussion related to the topics or issues of part 3, prompted by examiner questions to each of the examinees. On the face of it, the validity of the CAE is enhanced by the inclusion of a variety of tasks and interaction between participants instead of only between one participant and an examiner.

The assessment of speaking in the CAE has five analytical assessment scales: Grammatical Resource, Lexical Resource, Discourse Management, Pronunciation, and Interactive Communication.¹⁰ The Cambridge ESOL unit helped develop the CEFR, and the Cambridge exams map onto the CEFR levels.¹¹ The Pronunciation scale of the CAE emphasizes intelligibility at the lower levels (A1–B1), adding intonation, stress, and articulation at intermediate and advanced levels (B2–C2). According to the Pronunciation scale descriptors, intonation is to be assessed in terms of “appropriacy” while articulation and stress are to be assessed in terms of “accuracy,” which in the case of stress refers to its placement in words and sentences. Otherwise, the Pronunciation scale descriptors for B2–C2 are the same except that those for level B2 are all qualified by the word “generally” (e.g. C1: “Intonation is appropriate.” vs. B2: “Intonation is generally appropriate.”). In its descriptors, the CAE Discourse Management scale resembles the IELTS Fluency and Coherence scale in terms of its focus, while the Interactive Communication scale assesses the testee’s performance in the joint construction of communicative acts.

The Step 2 Clinical Skills (CS) Performance Test

Our final example of a test using human raters is the Step 2 Clinical Skills (CS) performance exam, which has been designed for a specific context and purpose and which includes measures of pronunciation skill or proficiency as part of certification for medical practice in the United States.

The Step 2 CS exam is part of the U. S. Medical Licensing Examination “required of all physicians who seek graduate training positions in the United States, including graduates from medical schools located in the United States and other countries” (van Zanten, 2011, p. 78). The exam evaluates candidates on the dimensions of Spoken English Proficiency, Communication and Interpersonal Skills, and Integrated Clinical Encounter, based on their interactions with actors simulating patients whom the candidate must interact with “as they would with actual patients, gathering relevant data through questioning the patients about their illnesses, performing physical exams, counseling patients on risky behaviors, and providing [them] with information about possible diagnosis and follow-up plans” (van Zanten, 2011, p. 79). Candidates perform 12 simulated interviews of up to 15 minutes each, and are given a further 10 minutes to write up a patient note from the interview. The raters of the test are the simulated patients (SPs), who undergo specific rater training for this purpose.

As van Zanten (2011) describes it, the Spoken English Proficiency rating is a “holistic judgment of...overall listener effort and how effectively the physician communicated, not simply whether or not the physician had a foreign accent” (p. 81), but with specific consideration of mispronunciations that might have caused difficulty in listener understanding or communication breakdown. The candidate’s Communication and Interpersonal Skills are measured in the dimensions of questioning skills, information-sharing skills, and professional manner and rapport (p. 80). Some aspects of pronunciation, especially prosody, may be evaluated indirectly in these three categories, as questioning skills includes the candidate’s “facilitating remarks” to the patient; information-sharing skills include “acknowledge[ing] patient issues and concerns and...then clearly respond[ing];” and professional manner and rapport includes “encourage[ing] additional questions or discussion, ...mak[ing] empathetic remarks, and...show[ing] consideration for patient comfort during the physical examination” (pp. 80–81).

The Step 2 CS exam is one of the most comprehensive and specific tests of oral proficiency that we are aware of, a model of thoroughness, specificity, and relevance in the design of tasks and assessment measures geared to sample a certain group in terms of their proficiency or competence in a range of behaviors involving spoken language. It has been extensively studied and validated according to traditional measures of validity (see van Zanten, 2011) and also has high validity of the wider sort that Messick (1996) and Winke (2011) have called for, including face and consequential validity. The simulation tasks and training procedures of this exam can serve as a model for the design of other kinds of specific-purposes oral language tests for jobs in which spoken language proficiency has a high impact, though the time required to take the test (approximately 7 hours), in addition to its physical requirements for the testing sites, the small number of test takers that can be accommodated in one administration, and the costs involved in developing and taking this test, would in many other work contexts not be feasible.

Some Widely Used Automated Systems for Oral Proficiency Assessment

Versant

One of the most extensively developed automated systems for oral proficiency assessment is Pearson Education's *Versant* technologies, originally developed as the *Phone Pass* system at the Ordinate company. The *Versant* system provides automated assessment by phone or on computer and computerized scoring for assessing spoken English, Aviation English (designed to assess the job-related spoken English proficiency of pilots and air traffic controllers), and spoken language assessments for Arabic, Chinese, Dutch, French, and Spanish.¹² According to the *Versant* website, which includes information about the form and the validation of the *Versant* tests (e.g., Bernstein, 2004; Suzuki et al., 2006), spoken proficiency is assessed by having test takers perform a variety of speaking tasks, including reading sentences aloud, repeating sentences, giving short answers to short questions, building sentences (by reordering phrases given out of order), retelling brief

stories, and supplying information or part of a conversation. The test thus aims to satisfy the requirement set by Riggensbach (1998) for task variety in spoken language assessment. The measures taken are Pronunciation, Fluency, Vocabulary, and Sentence Mastery. Scores are given in each of these categories on a 20–80 point scale and an overall score is calculated. Both the overall score and the individual test sections have been found to have high reliability, and the overall score correlates highly to human ratings and to other instruments used to assess spoken English (Bernstein, Van Moere, & Cheng, 2010), thus confirming criterion-related validity. The overall score has also been equated to the CEFR levels and hence to a number of other tests whose scores can be equated to the CEFR levels, such as IELTS and TOEFL tests.

The scoring is based on large databases of native and non-native speech, the first providing the language model against which a test taker's performance is assessed and the second providing language models for different levels of performance. The *Versant* system has proved effective in terms of its concurrent validity in comparison to human ratings and is more efficient and reliable (in terms of its criteria) than most tests requiring human administration and/or human rating. However, as pointed out in Chap. 5, the system (or earlier versions of it in Phone Pass) has been criticized for the lack of measurement of prosody and the limitations of assessing short samples of speech (Hincks, 2001). As it has been developed over time, its developers have attempted to address matters of content/construct validity by increasing the number and variety of tasks performed and by adding open-ended and collaborative tasks.

TOEFL iBT Speaking Test and Next-Generation Speaking Assessment by Machine

The Speaking section of the TOEFL online iBT¹³ incorporates a range of communicative tasks integrated with other language skills and using a range of native speaker accents in its recorded material. It has 6 tasks that include expressing an opinion on a familiar topic, speaking based on reading, and speaking based on listening. Performance is scored on four levels for Delivery, Language Use, and Topic Development using both human raters and automated scoring. Scoring rubrics¹⁴ are clear and consistent across the four rated levels of performance (above level 0):

Level 1: Consistent pronunciation and intonation problems cause considerable listener effort and frequently obscure meaning. Delivery is choppy, fragmented, or telegraphic. Speech contains frequent pauses and hesitations.

Level 2: Speech is basically intelligible, though listener effort is needed because of unclear articulation, awkward intonation, or choppy rhythm/pace; meaning may be obscured in places.

Level 3: Speech is generally clear, with some fluidity of expression, though minor difficulties with pronunciation, intonation, or pacing are noticeable and may require listener effort at times (though overall intelligibility is not significantly affected).

Level 4: Generally well-paced flow (fluid expression). Speech is clear. It may include minor lapses, or minor difficulties with pronunciation or intonation patterns, which do not affect overall intelligibility.

Research has determined a “satisfactory” level of reliability and “dependability of classification decisions made based on different cut scores” of the iBT Speaking section (Sawaki & Sinharay, 2013, Abstract). Other validation research (Ockey, Koyama, Setoguchi, & Sun, 2015) has shown strong correlations between the iBT Speaking scores and some facets of oral ability—pronunciation, fluency, and vocabulary/grammar—and moderate correlations with others—interactional competence, descriptive skill, and delivery skill. For more than a decade, ETS has been developing speech assessment geared to measuring spontaneous speech. Their website¹⁵ describes their next-generation *SpeechRater* software, in development since 2006 and being used to score the TOEFL Practice Online test, as an “advanced spoken-response scoring application targeted to score spontaneous responses, in which the range of valid responses is open-ended rather than narrowly determined by the item stimulus.” *SpeechRater* uses ASR technology “specially adapted for use with nonnative English,” along with natural language processing and speech-processing algorithms which “define a ‘profile’ of the [test taker’s] speech on a number of linguistic dimensions, including fluency, pronunciation, vocabulary usage, grammatical complexity and prosody. A model of speaking proficiency is then applied to these features in order to assign a final score to the response.” Part of the ETS development effort for *SpeechRater* is to include pragmatic competencies and discourse structure

across different task types within their model constructs and analyzable features. Research is ongoing into the different constructs, measured dimensions of speech, and model of speaking proficiency on which the *SpeechRater* system is based (e.g., Loukina, Zechner, Chen, & Heilman, 2015; Xi, Higgins, Zechner, & Williamson, 2012).

Pronunciation Aptitude

Other than some of the Versant test tasks, pronunciation is not usually assessed outside a classroom setting as an individual aspect of L2 proficiency or achievement. The separate assessment of pronunciation is however the focus of some of the subtests of language aptitude. The Modern Language Aptitude Test (MLAT; Carroll & Sapon, 2002/1959) taps into perceptual aspects of pronunciation in its first three subtests. According to a description and sample items from the test provided by the Language Learning and Testing Foundation¹⁶ the first part of the test is Number Learning, in which test takers learn a set of simple and easily pronounceable number words for an unknown language (e.g., /ba/ and /ti/) based on aural input and then use this explicit knowledge to recognize those numbers in different combinations. This subtest thus incorporates the ability to discriminate different sounds (an analytical ability underlying perception of speech as intelligible and all other spoken language abilities) and to retain these in memory.

The second part of the MLAT is Phonetic Script, in which test takers learn sets of four one-syllable nonsense words (e.g. *buk* and *geeb*) given in aural and written form using a made-up script (which has commonalities with IPA symbols and other systems of pronunciation symbols) and then must show recall by matching heard nonsense words to the correct script out of the four choices. This subtest, like the first one, incorporates the ability to discriminate different sounds which underlies speech perception as well as the ability to learn sound-symbol correspondences (an analytical ability underlying literacy) and to retain these in memory—which is presumably aided for literate learners by the link to written graphemes.

The third part of the MLA, Spelling Clues, gives altered spellings of English words based on their approximate pronunciation (e.g., *kloz* for

“clothes” and *restrnt* for “restraint”), though no aural versions are presented, and asks test takers to find the word with the right meaning from a list of five choices. It is interesting that this subtest is the one most often used from among the MLAT individual subtests as a measure of pronunciation aptitude, since it does not directly test either perception or production of pronunciation but rather directly tests English language orthographic and lexical knowledge.

The LLAMA Language Aptitude Test (Meara, 2005) includes two subtests related to pronunciation, LLAMA D and LLAMMA E. The LLAMA D subtest is based on a language which is expected to be unknown to almost all L2 learners, a language of Northwestern British Columbia, replacing the original version of the test based on Turkish, which turned out to be known to many of the test takers (Meara, 2005, p. 2). The 2005 version is a new type of subtest described as assessing “sound recognition” using names of flowers and other natural objects in the selected obscure Canadian Indian language that are presented in the form of synthesized sounds, that is, computer-generated speech. Test takers hear 10 of these synthesized words in the first phase of the subtest, and then are given further words which they have to indicate as the same or different from those they heard before. The LLAMA E subtest is similar to the MLAT subtest Phonetic Script but with less similarity to any phonetic spelling conventions. Test takers hear 22 recorded syllables and learn symbols for these by clicking on labeled buttons on a specially designed computer screen. Each sound is represented by a combination of a letter symbol and a numeral, such as “3e” or “9è.” In the test phase, recordings of 2-syllable sequences are heard with two choices presented for their spelling, such as “0è3è” or “3e9è” (Meara, 2005, pp. 11–13).

These tests of pronunciation aptitude are sometimes used to assess pronunciation proficiency in research studies (e.g., Granena & Long, 2013; Hu et al., 2013; Sáfár & Kormos, 2008; Saito, 2018), and these same measures, or others (e.g., based on imitation), might be usefully considered in standardized tests of speaking proficiency that include pronunciation measures, as a way to gain a view of a learner’s pronunciation proficiency separate from other aspects of speaking.

Classroom Assessment

The assessment of pronunciation in a classroom setting may address any of the different dimensions of pronunciation competence reviewed above for production (accuracy, fluency, intelligibility, or comprehensibility) or perception (segmental or prosodic), using one or more focused tasks (e.g., minimal pair discrimination, word or sentence repetition/imitation) or global tasks (e.g., picture description or verbal response to questions). Classroom assessment may address pronunciation in connection with other language skills (e.g., overall speaking skills or other aspects of communicative competence) or as an autonomous area of language, by means of tasks designed to factor out lexical or grammatical knowledge or any kind of specific language knowledge (e.g., by repetition or imitation of minimal pair syllables or nonsense words, as in some tests of pronunciation aptitude). As in all curriculum and classroom activities, student need and the specific focus and goals of instruction are crucial determinants of what and how to assess.

Taking into consideration the different facets of pronunciation and the different ways that it can be assessed, some general guidelines can be given for the assessment of pronunciation as part of classroom practice.

1. Assess different aspects of pronunciation:
 - perception, production; segmentals, prosody; intelligibility, fluency; referential and social meaning.
2. Assess pronunciation in a range of tasks and contexts:
 - at various levels of task difficulty;
 - in production and perception of speech;
 - in relation to written and read language;
 - in tasks when it is, and is not, the focus of the speaker's performance;
 - under conditions of slow and controlled or emphatic speech as well as fast and automatized performance;
 - in known and unknown words and when speaking on familiar or rehearsed topics and on unfamiliar or unrehearsed topics;

- while speaking under different audience conditions (interacting in a pair, participating in a small group, and giving a prepared speech in front of an audience).
3. Assess a student's ability to correct when given a specific model in the form of corrective feedback as well as to self-correct when not understood.
 4. Assess pronunciation in terms of the contribution it makes to communication in specific tasks.
 5. Assess the ability to purposefully adjust pronunciation as an aspect of speaking style and expression of identity and affiliation.

An approach to formative assessment or diagnosis of pronunciation problems is Pennington's (1996) "Pedagogical classification of pronunciation errors and problems" as:

1. Most important pronunciation errors or problems
 - (a) Those which occur most frequently
 - (b) Those which are the most serious—that is, which have the greatest effect on intelligibility
2. Those for which there is the greatest chance of successful remediation—that is, those which will be easiest to correct
3. Errors or problems identified by the learners themselves as aspects of their language needing attention
 - (a) Stereotyped errors
 - (b) Errors causing embarrassment or obvious miscommunication
 - (c) Items they would like to pronounce correctly
4. Errors or problems in areas of language that are of particular importance for the learner's social, academic, or professional needs
 - (a) Errors or problems in communicational pragmatics
 - (b) Errors or problems related to specialized areas of language (pp. 256–257)

Toward the Future of Pronunciation Assessment

We have come a long way since the 1950s in developing standardized tests of oral proficiency but still have a way to go in understanding what pronunciation is and how pronunciation is appropriately assessed. In the view of Trofimovich and Isaacs (2017), “more theory building is required to understand the nature of the phenomena being targeted through assessment and, specifically, to better understand major global constructs in L2 pronunciation so they can be better operationalized in assessment instruments” (p. 267). Research on L2 pronunciation is doing a good job of analyzing pronunciation and conducting “empirical evaluation of the meaning and consequences of measurement... combin[ing] scientific inquiry with rational argument” (Messick, 1996, p. 2) to consider the best ways to assess pronunciation for different purposes. As Messick (1996) made clear, “Because evidence is always incomplete, validation is essentially a matter of making the most reasonable case to guide both current use of [a] test and current research to advance understanding of what the test scores mean” (p. 1).

Trofimovich and Isaacs (2017) call for more research in a number of areas:

- Pronunciation assessment for languages other than English;
- Speech in non-academic contexts;
- Lingua franca L2 users;
- Sociolinguistic and pragmatic functions of L2 pronunciation;
- Incorporating nonnative pronunciation models and standards in assessments;
- Classroom-based assessment, including the instructional effectiveness of incidental form-focused instruction (i.e. corrective feedback) on L2 pronunciation development;
- Learners’ self-assessment of pronunciation, and ways to help learners calibrate their perceptions to those of their interlocutors;
- Testing instruments and procedures involving interpersonal interaction in authentic situations, including interactional paired and group assessments;

- Pronunciation features that account for communication breakdowns and lack of understanding;
- The role of holistic pronunciation-related constructs such as intelligibility in relation to more discrete L2 speech measures and to listener/rater/interlocutor variables. (pp. 266–267)

We would also like to see more attention to formative assessment using computerized learning and assessment programs showing the extent as well as the manner of divergence of a speaker's performance from that of a specified target—preferably one that the language learner specifies.

Harding (2017) makes a number of technical recommendations as regards the assessment of pronunciation:

1. Include all assessed elements of pronunciation across rating scale levels (segmental and suprasegmental features)...;
2. Avoid abstract terminology such as 'natural', which requires reference to a scale-external standard and which may function as an implicit normative concept;
3. Avoid incongruous references to other skill/knowledge areas (e.g. lexicogrammar) unless these are purposefully included across all levels...;
4. Maintain consistency of terminology across the scale to reduce the challenge for raters in following the trajectory of a feature across levels;
5. Keep scales brief—six levels appear to be sufficient. Within level descriptors, ... [t]hree to five clauses per level may be optimal.... (pp. 29–30)

In his construct recommendations, Harding (2017) suggests that test designers “[c]onsider collapsing pronunciation and fluency into the same criterion” (p. 30), as is done in the criterion labeled “Delivery,” while striving to keep ratings of oral language proficiency based on pronunciation and grammatical factors separate, such as by including descriptors that differentiate grammatical accuracy and pronunciation accuracy at all levels. We note however the option taken in the IELTS Speaking exam of separating the rating of pronunciation and fluency and grouping the latter with timing and organization of speech into a Fluency and Coherence

criterion. We suggest that the two ways of handling fluency—by grouping it together with pronunciation or with coherence—should be compared in future validation studies and theoretical considerations of the nature of pronunciation and speaking proficiency more generally. Harding further recommends “[removing] references to foreign accent in pronunciation scales unless there are clear purpose-driven reasons to assess strength of accent (rather than intelligibility)” (p. 30). In this connection, we note that considerations of accentedness are not necessarily of no consequence for pronunciation assessment, as they are relevant to perceptions and evaluations of speech and of speakers by naïve listeners, who may be speakers’ employers, coworkers, and customers, for example. Another of Harding’s recommendations is “use of eye-tracking methodology to fully understand how raters engage with specific scales: what they look at, when they look at it, what they don’t look at, and how variations in scale complexity, layout and terminology might affect the ways in which raters interact with rating scales” (p. 31). As a new approach to obtaining information about human ratings, eye-tracking is worth exploring in ongoing research efforts to understand and validate pronunciation scales and rating procedures.

Trofimovich and Isaacs (2017) conclude their review of the state of the art in pronunciation assessment, as reflected in the chapters in their edited collection, by listing a number of questions that remain open and pressing, and that call for further research and test development efforts. We wish to highlight, in particular, their call for more investigation of:

- perceptions of pronunciation in different types of context and by different listener stakeholder groups, and how these effect speakers’ communicative success in real-world settings;
- ways in which technological measurements can help validate listener perceptions of speech;
- ways to mitigate sources of low inter-rater reliability related to listener background (e.g., accent familiarity) in high-stakes assessments of L2 speech, such as through rater screening and training;
- multicompetence and whether it could form the basis of a standard for pronunciation assessment, drawing on descriptions of proficient multicompetent learners or test takers. (pp. 267–268)

We would add the need to further consider the ways in which pronunciation contributes to strategic language behavior or competence, “the way speakers use communicative resources to achieve their communicative goals, within the constraints of their knowledge and of the situation in which communication takes place” (Pennington, 2008, p. 164). Phonology is a key element of linguistic and communicative competence in spoken language that relates to listening comprehension and to written language as well (through the orthographic system). Such a concept of pronunciation goes far beyond that which is conceived in terms of either accuracy/native-ness or intelligibility, suggesting that assessment should not only be tied to tasks within specific contexts and purposes but also that pronunciation can suitably be assessed in terms of the extent to which a speaker is able to alter pronunciation under different circumstances, in order to achieve different goals. One of the potential limitations of L2 speech, including L2 lingua franca speech, is the degree of flexibility in speech style, which is an important concern of language learners as they advance in proficiency and use the L2 in more contexts with more different speakers. It is especially a concern of those who wish to develop a translingual or plurilingual identity or persona and to convey different aspects of their translingual or plurilingual identity or competence through style-shifting, or specifically what might be referred to as “pronunciation styling.” With the notable exception of IELTS, limited flexibility in L2 performance has hardly been recognized by the speech testing community, much less the subgroup of testers focused on pronunciation. The expression of identity and affiliation through pronunciation and how this might be assessed is a consideration that has received little attention to date.

Thus, we see a need in pronunciation assessment to develop social perception and production tasks involving considerations of context and audience that will incorporate pronunciation pragmatics, such as perception and production of contextualization cues to metamessages (Chap. 1) and a shifting focus on different aspects of a message, its meaning or its form, in response to audience feedback, thereby incorporating aspects of pronunciation proficiency needed for interactional competence (Young, 2000) and reciprocity (Bygate, 1987). This broader focus on pronunciation in social context will also address perception and expression of identity affiliation, accommodation, and rapport.

Balancing this need at what may be considered the “macro” level of pronunciation is a need at the other end of the pronunciation spectrum, the “micro” level of auditory perception and articulatory production that is foundational for intelligibility and all higher levels of linguistic performance. At the micro level, we recognize a need for more attention to pronunciation considered narrowly as an autonomous area of language, in both:

- *its auditory perceptual aspect*, such as the ability to recognize cues to meaningful contrasts in pronunciation at the segmental and prosodic level; and
- *its productive aspect*, such as the ability to accurately repeat or imitate nonsense words (e.g., as in some parts of the MLAT or LLAMA and also in the research on individual differences in L2 proficiency and ultimate attainment) and real language (e.g., sentences, as in the Versant computer-based test and in research on individual differences in L2 proficiency and ultimate attainment reviewed in Chap. 2).

Combining the macro and micro levels of pronunciation competence, there is a need for more research into assessing listeners’ perception and production of the types of segmental and prosodic features and contrasts that signal interpersonal as well as lexicogrammatical meaning. The assessment of communicative performance in call centers is clearly an area for future research and development at both the macro and micro levels of pronunciation. The ability to establish mutual understanding and positive interpersonal dynamics between L2-speaking customer service representatives (CSRs) and their overseas callers, as a basis for effective communication, is the focal concern in outsourced call centers. Davies (2010) has identified the following areas as those requiring attention in future performance assessment in the outsourced overseas call center context:

- clarity (“the form of what is said not the content”);
- intelligibility (“the content of what is said”); and
- rapport (“being on the same wavelength as the person with whom you are talking”), which is closely related to empathy. (p. 246)

By identifying these aspects of communicative performance, all of which intersect with pronunciation in its segmental and prosodic aspects, as the needed focus of assessment and further research in call centers, Davies (2010) has outlined a course for the future that ties together research and practice in pronunciation with the real-world communicative needs of L2 speakers. In so doing, he has perhaps given direction for other practitioners with an interest in pronunciation to tie assessment to communicative performance.

Concluding Remarks

This chapter has considered language assessment in respect of the complex view of pronunciation presented in Chap. 1. Most of the attention has been trained on the ways in which pronunciation is assessed within speaking proficiency more generally, as this has usually been the approach taken in standardized tests and also in some contexts of practice, notably, in many language classrooms. We have considered the nature of pronunciation and how testers' understanding of the nature and construct of pronunciation impacts the reliability and validity of different approaches to assessment. For the most part, language testers are not paying sufficient attention to pronunciation, what it is, how it is measured and rated, and the validation of pronunciation ratings in assessment instruments. A comparison of human and machine ratings has pointed up the benefits and drawbacks of each of these ways of assessing pronunciation and the need to continue refining the techniques used in these different approaches.

In continuing to scrutinize and revise pronunciation assessments through ongoing validation studies and other research, pronunciation and testing specialists will continue to improve those assessments even as they continue to interrogate the construct of pronunciation and what competence in this area of language and communication includes. In so doing, they will be impacting the lives of the huge worldwide community of L2 learners and their opportunities for further study and employment related to their assessed language abilities, even as they are also contributing to improving understanding of this pervasive surface aspect of language, the many kinds of deeper level linguistic properties which it enacts, and the meanings which it signifies.

Notes

1. Computer-based testing programs that measure how far a speaker's performance diverges from a set norm could perhaps be adapted to both summative and formative assessment of pronunciation in such attempts at "dialect adjustment."
2. This is a vexed concept, as discussed in Chap. 1, that recognizes an accent only in nonstandard or L2 speech and does not accord with the reality that all speech is accented.
3. <http://www.actfl.org/publications/guidelines-and-manuals/actfl-proficiency-guidelines-2012/english/speaking>
4. <http://www.govtilr.org/skills/ILRscale2.htm>
5. <http://www.govtilr.org/FAQ.htm#14>
6. A full report on the phonological scale revision process is available from the CEFR <https://rm.coe.int/phonological-scale-revision-process-report-cefr/168073fff9>
7. <http://www.ielts.org/what-is-ielts/ielts-introduction>
8. <http://www.ielts.org/-/media/pdfs/speaking-band-descriptors.ashx?la=en>
9. <http://www.cambridgeenglish.org/exams/advanced/exam-format>
10. <http://www.cambridgeenglish.org/images/167804-cambridge-english-advanced-handbook.pdf>
11. <https://www.examenglish.com/examscomparison.php>
12. <https://www.versanttest.com>
13. <http://www.ets.org/toefl/ibt/about/content>
14. http://www.ets.org/s/toefl/pdf/toefl_speaking_rubrics.pdf
15. http://www.ets.org/research/topics/as_nlp/speech
16. http://lftf.net/wp-content/uploads/2011/09/Test_mlat.pdf

References

- Abrahamsson, N., & Hyldenstam, K. (2009). Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language Learning*, 59(2), 249–306. <https://doi.org/10.1111/j.1467-9922.2009.00507.x>
- American Council on the Teaching of Foreign Languages. (2012). *ACTFL proficiency guidelines 2012*. Alexandria, VA: American Council on the Teaching of Foreign Languages.

- Anderson-Hsieh, J., & Koehler, K. (1988). The effect of foreign accent and speaking rate on native speaker comprehension. *Language Learning*, 38(4), 561–613. <https://doi.org/10.1111/j.1467-1770.1988.tb00167.x>
- Ballard, L., & Winke, P. (2017). Students' attitudes towards English teachers' accents: The interplay of accent familiarity, comprehensibility, intelligibility, perceived native speaker status, and acceptability as a teacher. In T. Isaacs & P. Trofimovich (Eds.), *Second language pronunciation assessment: Interdisciplinary perspectives* (pp. 121–140). Bristol, UK and Blue Ridge Summit, PA: Multilingual Matters.
- Bernstein, J. (2004). Spoken language testing: Ordinate's SET-10. Presentation given at the *TESOL Convention*, April 2004, Long Beach, CA. Retrieved September 2, 2017, from <https://www.versanttest.com/technology/featured/3.TESOL-04.SET.JB.pdf>
- Bernstein, J., Van Moere, A., & Cheng, J. (2010). Validating automated speaking tests. *Language Testing*, 27(3), 355–377. <https://doi.org/10.1177/0265532210364404>
- Bosker, H. R., Pinget, A.-F., Quené, H., Sanders, T., & de Jong, N. H. (2012). What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing*, 30(2), 159–175. <https://doi.org/10.1177/0265532212455394>
- Brown, J. D. (2005). *Testing in language programs: A comprehensive guide to English language assessment*. Upper Saddle River, NJ: Prentice Hall Regents.
- Browne, K., & Fulcher, G. (2017). Pronunciation and intelligibility in assessing spoken fluency. In T. Isaacs & P. Trofimovich (Eds.), *Second language pronunciation assessment: Interdisciplinary perspectives* (pp. 37–53). Bristol, UK and Blue Ridge Summit, PA: Multilingual Matters.
- Bygate, M. (1987). *Speaking*. Oxford: Oxford University Press.
- Carey, M. D., Mannell, R. H., & Dunn, P. K. (2011). Does a rater's familiarity with a candidate's pronunciation affect the rating in oral proficiency interviews? *Language Testing*, 28(2), 201–219. <https://doi.org/10.1177/0265532210393704>
- Carroll, J. B., & Sapon, S. (2002/1959). *Modern language aptitude test*. San Antonio, TX and Bethesda, MD: Psychological Corporation and Second Language Teaching, Inc.
- Chalhoub-Deville, M. (2003). Second language interaction: Current perspectives and future trends. *Language Testing*, 20(4), 369–383. <https://doi.org/10.1191/0265532203lt264oa>
- Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.

- Council of Europe. (2017). Common European framework of reference for languages: Learning, teaching, assessment. Companion volume with new descriptors. Retrieved February 26, 2018, from <https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989>
- 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. <https://doi.org/10.1080/09658416.2011.570347>
- Couper, G. (2015). Applying theories of language and learning to teaching pronunciation. In M. Reed & J. M. Levis (Eds.), *Handbook of English pronunciation* (pp. 413–432). New York: John Wiley & Sons.
- Cruttenden, A. (2014). *Gimson's pronunciation of English* (8th ed.). London: Routledge.
- Cucchiari, C., Strik, H., Binnenpoorte, D., & Boves, L. (2002). Pronunciation evaluation in read and spontaneous speech: A comparison between human ratings and automatic scores. In A. James & J. Leather (Eds.), *New sounds 2000, Proceedings of the fourth international symposium on the Acquisition of Second-Language Speech*, University of Amsterdam, September 2000 (pp. 72–79). Klagenfurt, Austria: University of Austria.
- Cucchiari, C., Strik, H., & Boves, L. (2000). Quantitative assessment of second language learners' fluency by means of automatic speech recognition technology. *Journal of the Acoustical Society of America*, 107(2), 989–999. <https://doi.org/10.1121/1.428279>
- Davies, A. (2010). Language assessment in call centres: The case of the customer service representative. In G. Forey & J. Lockwood (Eds.), *Globalization, communication and the workplace* (pp. 242–248). London: Continuum.
- Davies, A. (2017). Commentary on the native speaker status in pronunciation research. In T. Isaacs & P. Trofimovich (Eds.), *Second language pronunciation assessment: Interdisciplinary perspectives* (pp. 185–192). Bristol, UK and Blue Ridge Summit, PA: Multilingual Matters.
- de Bot, K. (1992). A bilingual production model: Levelt's 'speaking' model adapted. *Applied Linguistics*, 13(1), 1–24. Retrieved January 10, 2018, from https://www.researchgate.net/profile/Kees_De_Bot2/publication/249237191_A_Bilingual_Production_Model_Levelt%27s_%27Speaking%27_Model_Adapted/links/0deec537b3b8b45d98000000/A-Bilingual-Production-Model-Levels-Speaking-Model-Adapted.pdf
- Derwing, T. M., Rossiter, M. J., Munro, M. J., & Thomson, R. I. (2004). Second language fluency: Judgments on different tasks. *Language Learning*, 54(4), 655–679. <https://doi.org/10.1111/j.1467-9922.2004.00282.x>

- Fulcher, G. (1996). Testing tasks: Issues in task design and the group oral. *Language Testing*, 13(1), 23–51. <https://doi.org/10.1177/026553229601300103>
- Fulcher, G. (2015). Assessing second language speaking. *Language Teaching*, 48(2), 198–216. <https://doi.org/10.1017/S0261444814000391>
- Granena, G., & Long, M. H. (2013). Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research*, 29(3), 311–343. <https://doi.org/10.1177/0267658312461497>
- Harding, L. (2017). What do raters need in a pronunciation scale? The user's view. In T. Isaacs & P. Trofimovich (Eds.), *Second language pronunciation assessment: Interdisciplinary perspectives* (pp. 12–34). Bristol, UK and Blue Ridge Summit, PA: Multilingual Matters.
- Hincks, R. (2001). Using speech recognition to evaluate skills in spoken English. Lund University, Dept. of Linguistics *Working Papers*, 49, 58–61. Retrieved August 15, 2017, from journals.lub.lu.se/index.php/LWPL/article/viewFile/2367/1942
- Hu, X., Ackermann, J., Martin, J. A., Erb, M., Winkler, S., & Reiterer, S. M. (2013). Language aptitude for pronunciation in advanced second language (L2) learners: Behavioural predictors and neural substrates. *Brain and Language*, 127(3), 366–376. <https://doi.org/10.1016/j.bandl.2012.11.006>
- Isaacs, T. (2014). Assessing pronunciation. In A. J. Kunnan (Ed.), *The companion to language assessment* (pp. 140–155). Hoboken, NJ: Wiley-Blackwell.
- Isaacs, T., & Trofimovich, P. (2012). Deconstructing comprehensibility. *Studies in Second Language Acquisition*, 34(4), 475–505. <https://doi.org/10.1017/S0272263112000150>
- Isaacs, T., & Trofimovich, P. (2017). Key themes, constructs and interdisciplinary perspectives in second language pronunciation assessment. In T. Isaacs & P. Trofimovich (Eds.), *Second language pronunciation assessment: Interdisciplinary perspectives* (pp. 3–11). Bristol, UK and Blue Ridge Summit, PA: Multilingual Matters.
- Isaacs, T., Trofimovich, P., Yu, G., & Chereau, B. M. (2011). Examining the linguistic aspects of speech that most efficiently discriminate between upper levels of the revised IELTS pronunciation scale. *IELTS Research Reports Online*, 4, 1–48. Retrieved September 2, 2017, from http://www.ielts.org/-/media/research-reports/ielts_online_rr_2015-4.ashx
- Jenkins, J. (2000). *The phonology of English as an international language*. Oxford: Oxford University Press.
- Johnson, M. (2000). Interaction in the oral proficiency interview: Problems of validity. *Pragmatics*, 10(2), 215–231. <https://doi.org/10.1075/prag.10.2.03joh>

- Kane, M. T. (1990). An argument-based approach to validation. Research report no. ACT-RR-90-13. Iowa City, Iowa: American College Testing Program. ERIC Document No. ED 336428.
- Kang, O., Moran, M., & Vo, S. (2016). Perceptual judgments of accented speech by listeners from different first language backgrounds. *TESL-EJ*, 20(1). Retrieved February 22, 2018, from <http://www.tesl-ej.org/wordpress/issues/volume20/ej77/ej77a1/>
- Kormos, J. (2006). *Speech production and second language acquisition*. Mahwah, NJ: Lawrence Erlbaum.
- Labov, W. (1966). *The social stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.
- Lai, E. R. (2011, May). Performance assessment: Some new thoughts on an old idea. *Pearson Bulletin*, 20. Retrieved December 12, 2016, from <https://www.pearsonassessments.com/research>
- Lennon, P. (1990). Investigating fluency in EFL: A qualitative approach. *Language Learning*, 40(3), 387–417. <https://doi.org/10.1111/j.1467-1770.1990.tb00669.x>
- Levis, J. M. (2006). Pronunciation and the assessment of spoken language. In R. Hughes (Ed.), *Spoken English, TESOL and applied linguistics: Challenges for theory and practice* (pp. 245–270). New York: Palgrave Macmillan.
- Levis, J. (2007). Computer technology in teaching and researching pronunciation. *Annual Review of Applied Linguistics*, 27, 184–202. <https://doi.org/10.1017/S0267190508070098>
- Loukina, A., Zechner, K., Chen, L., & Heilman, M. (2015). Feature selection for automated speech scoring. In *Proceedings of the tenth workshop on innovative use of NLP for building educational applications* (pp. 12–19). Denver, CO: Association for Computational Linguistics.
- Meara, P. (2005). *LLAMA language aptitude tests*. Swansea, UK: Lognostics.
- Messick, S. (1996). Validity of performance assessments. In G. W. Phillips (Ed.), *Technical Issues in Large-Scale Performance Assessments* (Chap. 1, pp. 1–18). Washington, DC: National Center for Education Statistics, U. S. Department of Education.
- Mompean, J. A. (2006). Introduction: Cognitive phonology in cognitive linguistics. *International Journal of English Studies*, 6, vii–xii. Retrieved August 17, 2017, from <https://www.scribd.com/document/298249578/Cognitive-Phonology-Ijes-6-2>
- Mompean, J. A. (2014). Phonology. In J. R. Taylor & J. Littlemore (Eds.), *Bloomsbury companion to cognitive linguistics* (pp. 357–392). London: Bloomsbury.

- Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 45(1), 285–310. <https://doi.org/10.1111/j.1467-1770.1995.tb00963.x>
- Munro, M. J., & Derwing, T. M. (1998). The effects of speaking rate on the comprehensibility of native and foreign accented speech. *Language Learning*, 48(2), 159–182. Retrieved January 10, 2018, from <http://psycnet.apa.org/doi/10.1111/1467-9922.00038>
- Munro, M. J., & Derwing, T. M. (2001). Modeling perceptions of the accentedness and comprehensibility of L2 speech: The role of speaking rate. *Studies in Second Language Acquisition*, 23(4), 451–468. Retrieved January 10, 2018, from <https://www.cambridge.org/core/journals/studies-in-second-language-acquisition/article/modeling-perceptions-of-the-accentedness-and-comprehensibility-of-l2-speech-the-role-of-speaking-rate/E22C75E244694BF6D6E083149FBFEFA51>
- Munro, M. J., Derwing, T. M., & Morton, S. L. (2006). The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition*, 28(1), 111–131. <https://doi.org/10.1017/S0272263106060049>
- Ockey, G. J., Koyama, D., Setoguchi, E., & Sun, A. (2015). The extent to which TOEFL iBT speaking scores are associated with performance on oral language tasks and oral ability components for Japanese university students. *Language Testing*, 32(1), 39–62. <https://doi.org/10.1177/0265532214538014>
- Pennington, M. C. (1989). Teaching pronunciation from the top down. *RELC Journal*, 20(1), 20–38. <https://doi.org/10.1177/003368828902000103>
- Pennington, M. C. (1990). The context of L2 phonology. In H. Burmeister & P. Rounds (Eds.), *Variability in second language acquisition: Proceedings of the tenth meeting of the second language research forum* (pp. 541–564). Eugene, OR: University of Oregon. Available at https://books.google.com/books/about/Variability_in_second_language_acquisiti.html?id=wkYqAQAAIAAJ
- Pennington, M. C. (1992). Discourse factors in second language phonology: An exploratory study. In J. Leather & A. James (Eds.), *New sounds 92, Proceedings of the 1992 Amsterdam Symposium on the Acquisition of Second-Language Speech*, University of Amsterdam, April 1992 (pp. 137–155). Amsterdam: University of Amsterdam.
- Pennington, M. C. (1996). *Phonology in English language teaching: An international approach*. London and New York: Longman.
- Pennington, M. C. (2008). A strategic view of pronunciation in English as a second language. In G. Cane (Ed.), *Strategies in language learning and teaching* (pp. 104–115). Anthology Series 49. Singapore: SEAMEO Regional Language Centre.

- Pennington, M. C. (2015). Research, theory, and practice in second language phonology: A review and directions for the future. In J. A. Mompean & J. Fouz-González (Eds.), *Investigating English pronunciation: Trends and directions* (pp. 149–173). Basingstoke, UK and New York, NY: Palgrave Macmillan.
- Piccardo, E. (2016) *Common European Framework of Reference for Languages: Learning, teaching, assessment phonological scale revision process report*. Retrieved February 26, 2018, from <https://rm.coe.int/phonological-scale-revision-process-report-cefr/168073fff9>
- Pinget, A.-F., Bosker, H. R., Quené, H., & de Jong, N. H. (2014). Native speakers' perceptions of fluency and accent in L2 speech. *Language Testing*, 31(3), 349–365. <https://doi.org/10.1177/0265532214526177>
- Poulisse, N. (1999). *Slips of the tongue: Speech errors in first and second language production*. Amsterdam: John Benjamins.
- Riggenbach, H. (1998). Evaluating learner interaction skills: Conversation at the micro level. In R. F. Young & A. W. He (Eds.), *Talking and testing: Discourse approaches to the assessment of oral proficiency* (pp. 53–67). Amsterdam: John Benjamins.
- Sáfar, A., & Kormos, J. (2008). Revisiting problems with foreign language aptitude. *IRAL—International Review of Applied Linguistics in Language Teaching*, 46(2), 113–136. <https://doi.org/10.1515/IRAL.2008.005>
- Saito, K. (2018). Effects of sound, vocabulary and grammar learning aptitude on adult second language speech attainment in foreign language classrooms. *Language Learning*, 67. <https://doi.org/10.1111/lang.12244>. Retrieved September 24, 2017, from Prepublication copy http://www.researchgate.net/publication/313337402_Effects_of_Sound_Vocabulary_and_Grammar_Learning_Aptitude_on_Adult_Second_Language_Speech_Attainment_in_Foreign_Language_Classrooms
- Saito, K., Trofimovich, P., & Isaacs, T. (2016). Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels. *Applied Psycholinguistics*, 37(2), 217–240. <https://doi.org/10.1017/S0142716414000502>
- Saito, K., Trofimovich, P., & Isaacs, T. (2017). Using listener judgements to investigate linguistic influences on L2 comprehensibility and accentedness: A validation and generalization study. *Applied Linguistics*, 38(4), 439–462. <https://doi.org/10.1093/applin/amv047>
- Sawaki, Y., & Sinharay, S. (2013). Investigating the value of TOEFL iBT section scores. *ETS Research Report RR-13-35*, TOEFLiBT-21.

- Scharenborg, O. (2007). Reaching over the gap: A review of efforts to link human and automatic speech recognition research. *Speech Communication*, 49(5), 336–347. <https://doi.org/10.1016/j.specom.2007.01.009>
- Segalowitz, N. (2010). *Cognitive bases of second language fluency*. New York: Routledge.
- Shavelson, R. J., Baxter, G. P., & Gao, X. (1993). Sampling variability of performance assessments. *Journal of Educational Measurement*, 30(3), 215–232. <https://doi.org/10.1111/j.1745-3984.1993.tb00424.x>
- Skehan, P. (2001). Tasks and language performance assessment. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching, and testing* (pp. 167–185). London: Longman.
- Skehan, P. (2007). Tradeoff and cognition: Two hypotheses regarding attention during task-based performance. Plenary delivered at the *2nd Annual Conference on Task-based Language Teaching*, University of Hawaii, September 2007. Retrieved from <http://www.hawaii.edu/tblt2007/PP/Plenaries/Skehan.ppt>
- Skehan, P. (2014). Task-based performance and task-based instruction: Research contributions. Plenary delivered at the *JALT Task-based Learning Special Interest Group meeting*, Osaka, Japan, May 2014. Retrieved from <http://www.tbtsig.org/wp-content/uploads/2014/05/Skehan-Osaka-May-2014.pdf>
- Smith, L., & Nelson, C. (1985). International intelligibility of English: Directions and resources. *World Englishes*, 4(3), 333–342. <https://doi.org/10.1111/j.1467-971X.1985.tb00423.x>
- Suzuki, M., Balogh, J., & Bernstein, J. (2006). *Psycholinguistics in computerized language testing*. Paper presented at the AAAL/CAAL conference, June 20–26, Montreal. Retrieved December 1, 2017, from <https://www.versanttest.com/technology/featured/6.AAAL2006.pdf>
- Tarone, E. (1998). Research on interlanguage variation: Implications for language testing. In L. F. Bachman & A. D. Cohen (Eds.), *Interfaces between second language acquisition and language testing research* (pp. 71–111). New York: Cambridge University Press.
- Tarone, E., Bigelow, M., & Hansen, K. (2009). *Literacy and second language oracy*. Oxford and New York: Oxford University Press.
- Trofimovich, P., & Isaacs, T. (2017). Second language pronunciation assessment: A look at the present and the future. In T. Isaacs & P. Trofimovich (Eds.), *Second language pronunciation assessment: Interdisciplinary perspectives* (pp. 259–266). Bristol, UK and Blue Ridge Summit, PA: Multilingual Matters.
- van Doremalen, J., Cuccharini, C., & Strik, H. (2010). Phoneme errors in read and spontaneous non-native speech: Relevance for CAPT system development. Interspeech Satellite Workshop. Second Language Studies: Acquisition,

- Learning, Education and Technology, Tokyo, Japan: Waseda University. Retrieved July 12, 2017, from http://www.gavo.t.u-tokyo.ac.jp/L2WS2010/papers/L2WS2010_O3-04.pdf
- van Lier, L. (1989). Reeling, writhing, drawing, stretching and fainting in coils: Oral proficiency interviews as conversation. *TESOL Quarterly*, 23(3), 489–508. <https://doi.org/10.2307/3586922>
- van Zanten, M. (2011). Evaluating the spoken English proficiency of international medical graduates for certification and licensure in the United States. In B. J. Hoekje & S. M. Tipton (Eds.), *English language and the medical profession: Instructing and assessing the communication skills of international physicians* (pp. 75–90). Bingley, UK; Leiden and Boston: Emerald and Brill.
- Wennerstrom, A. (2000). The role of intonation in second language fluency. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 102–127). Ann Arbor, MI: The University of Michigan Press.
- Winke, P. (2011). Evaluating the validity of a high-stakes ESL test: Why teachers' perceptions matter. *TESOL Quarterly*, 45(4), 628–660. <https://doi.org/10.5054/tq.2011.268063>
- Winke, P., Gass, S., & Myford, C. (2012). Raters' L2 background as a potential source of bias in rating oral performance. *Language Testing*, 30(2), 231–252. <https://doi.org/10.1177/0265532212456968>
- Xi, X., Higgins, D., Zechner, K., & Williamson, D. (2012). A comparison of two scoring methods for an automated speech scoring system. *Language Testing*, 29(3), 371–394. <https://doi.org/10.1177/0265532211425673>
- Yates, L., Zielinski, E., and Pryor, E. (2011). The assessment of pronunciation and the new IELTS Pronunciation Scale. In *IELTS Research Reports*, 12 (pp. 23–68). Melbourne & Manchester: IDP IELTS Australia and British Council.
- Young, R. F. (2000). Interactional competence: Challenges for validity. Paper presented at a joint symposium on Interdisciplinary Interfaces with Language Testing, held at the annual meeting of the American Association for Applied Linguistics and the Language Testing Research Colloquium, March 2000, Vancouver. Retrieved December 1, 2017, from http://www.english.wisc.edu/rfyoung/IC_C4V.Paper.PDF
- Young, R. F., & Milanovic, M. (1992). Discourse variation in oral proficiency interviews. *Studies in Second Language Acquisition*, 14(4), 403–424. <https://doi.org/10.1017/S0272263100011207>



7

Beyond the Language Classroom: Wider Applications of Pronunciation Research and Practice

Introduction

As societies, communities, and workplaces become increasingly diverse both linguistically and culturally, successful communication requires tolerance and a willingness to accommodate differences in experience, expectations, and skills. Communication difficulties can have many sources, including linguistic, cultural, and ideological differences, and pronunciation frequently plays a key role. Intelligibility is obviously a key component of successful communication and, as we stressed in Chap. 1, a two-way process involving both speaker and listener. Ensuring intelligibility in increasingly multilingual and multicultural encounters is a common challenge in many everyday social contexts, such as school classrooms, medical appointments, and call center exchanges. However, pronunciation is important not only for mutual intelligibility. The broader social role of language, and of pronunciation specifically, has been one of the key threads running through this book.

As introduced in Chap. 1, the way that people speak influences how others see them, as a central aspect of how they present themselves socially, both intentionally and unintentionally, in displaying their identity and

positioning themselves in specific communicative contexts; and people's social background influences both how they speak and also how they perceive speech (i.e., in the role of listener). Perhaps more than any other aspect of language, pronunciation can have a crucial impact not only on evaluations of an individual's communicative ability but also on the person's social standing, in terms of identity and status. Through the prosodic and segmental features of their pronunciation, speakers convey a great deal about their social class and values, their emotional state, and their personality. As a result, pronunciation can have a considerable impact on a person's life.

Due to its wide-ranging social role, the importance of pronunciation has been recognized in a diverse array of disciplines, not only linguistics but also education, psychology, sociology, management, and speech science; and each of these fields has developed its own research and applications relating to pronunciation. In the following sections, we consider this wider context of pronunciation research and practice beyond language teaching, in terms of some important applications within a broad range of disciplines, including in business and health communications, politics, forensic linguistics, speech technology, phonological impairment, and L1 literacy.

Professional Communication in an Increasingly Globalized and Diversified World

According to a recent report, English is now spoken to a useful level by approximately 1.75 billion people, that is, a quarter of the world's population (British Council, 2013). English is generally recognized as the language of global business, with many multinational corporations (e.g., Airbus, Daimler-Chrysler, Nokia, Renault, Samsung, Nestle, and Microsoft) adopting English as their common corporate language, in an attempt to facilitate international communication and economic advantage. English is also the most commonly used lingua franca in other multilingual professional contexts, such as diplomacy, research, and education. For example, in Europe, English is increasingly becoming the medium of teaching in many universities (Jenkins, 2014; Marsh & Laitinen, 2005).

While many see the pragmatic necessity of a lingua franca in such multilingual contexts, there is considerable concern and debate about the dominance of English, raising issues of linguistic, social, and political equality (Brutt-Griffler, 2008; Phillipson, 2008) alongside issues of intelligibility, communication breakdown, and prejudice.

While English may play an essential role as a lingua franca in many multilingual professional settings, it can also pose challenges, especially in situations where the majority of participants are L2 English speakers (Seidlhofer, 2003). A key concern is how to ensure linguistic equality and achieve international intelligibility and otherwise effective communication in such lingua franca contexts. In international events and encounters such as international conferences, meetings, and phone calls, effective oral communication—including pronunciation—is essential for managing transactions and interactions with clients, and for creating and maintaining professional relations with colleagues. Research into English as a lingua franca (ELF) has demonstrated the critical part played by pronunciation in maintaining successful communication between speakers of English with different first languages. As Jenkins (2000) notes, “There is evidence that phonological problems regularly get in the way of successful communication in international contexts” (p. 78). Such communication difficulties potentially result in lack of participation or misinterpretation between individuals or groups (Rogerson-Revell, 2008).

Other research is looking at communication in business contexts in which L1 English speakers are interacting with L2 English speakers at a considerable cultural and geographical distance, such as through overseas call centers (e.g., Forey & Lockwood, 2010), showing the kinds of difficulties that employees experience in these international intercultural encounters. Pronunciation plays an important role as well in intercultural encounters that take place between people living in the same country but who have different linguistic and cultural backgrounds, such as interactions between doctors and patients who do not share the same language and culture (e.g., Hoekje & Tipton, 2011; Roberts, Moss, Wass, Sarangi, & Jones, 2005). Pronunciation, both segmental phonology and prosody, is a factor in the effectiveness or ineffectiveness of communication in these international and intercultural settings, as in other communicative contexts where intelligibility and impact, together with status

and relationship, are in play. These attributes of communication which are contributed by features of pronunciation likewise give accent and prosody a central role in public speaking and specifically in politics (Black, 2016; Chozick, 2016).

Business and Professional Organizations

Accent in Business Contexts

In business contexts, such as sales or advertising, customers' evaluation of staff based on accent may shape consumer attitudes towards a company and its products or services. For the last two decades, research has been addressing these effects in certain areas, such as direct selling (DeShields & De los Santos, 2000) and advertising (Lalwani, Lewin, & Li, 2005; Reinares-Lara, Martín-Santana, & Muela-Molina, 2016).

Hoffman and Mai (2014, p. 140) suggest that accent can influence business interactions in three ways. First, the "social identity" effect describes how consumers assign accented speakers to social groups. Second, the "stereotyping" effect shows how accent impacts the evaluation of the speaker and the company. Third, the "speech processing" effect suggests that accent influences the processing and evaluation of the speaker's message. Hoffman and Mai (2014) claim that:

The extant studies predominantly demonstrate that foreign accented advertising spokespersons, job applicants or salespersons are evaluated, and often discriminated against, in terms of competence, social attractiveness and integrity.

Furthermore, a foreign accent decreases comprehensibility and therefore hampers message processing. In consequence, consumers express lower intentions to purchase and are less willing to recommend a product, service or brand than when interacting with a standard accented person. (p. 152)

While acknowledging that accent per se is not a barrier to communication, Hoffman and Mai (2014) do recommend that employers recognize the potential for accent to create customer bias and advise speech training for employees with "very strong regional and foreign accents" (p. 152).

However, they also offer a reminder that foreign accents can be advantageous in business communication, particularly if accent is used as a joint social identity marker in interaction to create in-group bonding.

International Business Meetings

Rogerson-Revell's (2008, 2010, 2014) discourse-based research into meetings in international in a large European organization describes some of the difficulties relating to spoken language. The research resulted from a concern within the European organization that an "oral language barrier" was causing L2 English speakers not to play as active a role in international organizational events, such as meetings, as L1 English speakers. As members of the organization stated, "The language problem is an important one to solve because, if we cannot communicate, then how can we expect to work together?" (Rogerson-Revell, 2008, p. 342).

Findings from a corpus of meetings data revealed a wide range of pronunciation proficiency among the organization's L2 English speakers, instances of unintelligibility, and various accommodation strategies that participants used to overcome such issues. One of the findings was that "phonological fluency is not necessarily related to accent and that accent is not necessarily problematic" (Rogerson-Revell, 2014, p. 154). There are examples in the meetings data where a strong or unfamiliar accent appears to impede intelligibility, as in this extract (Example 7.1) of an L1 Spanish speaker, "Felix." Prosodically prominent words are in boldface, unintelligible sections are marked with ??, and each dot indicates a pause of half a second:

Example 7.1. Extract from L1 Spanish Speaker (Felix)

a'nother [ə'nɒðə] 'question ... was a'round[ə] the **distri'bution** [dɪstrɪbʊʃən] .. in '**Spain** .. er ...?? I don(t) 'know ... his '**name** [hɪs naɪm] but in '**Spain** is 'operator de . ?? . **seguro**?? '**maybe** [maɪbi] .. I don't [t] 'know .. and so .. the ?? **distri'bution** .. 'tries to '**limit** [li:mi:t] . to **limit** .. the the '**scope** [skɒp]?? . and '**maybe** [maɪbi] .. be'cause they 'have [χæf] the '**main** [maɪn] '**bulk** [bulk] of '**business** ['buzɪnəs] but 'not the po'litical '**power**

(Rogerson-Revell, 2014, p. 147)

However, there are also examples in this data in which a speaker with a noticeable accent is phonologically fluent, particularly in terms of the use of accentuation and phrasing, as with “Jean” in the following extract (Example 7.2):

Example 7.2 Extract from L1 French Speaker (Jean)

'just to be 'sure 'all of you are a'ware .. we are 'holding the con'gress [kɔŋ'gre] in Pa'ris [pæ'ri:] in . er . 'June 'also . er .. we have pre'pared for 'you a ve'ry spe'cial pro'gramme .. er we will 'offer 'you expedi'tion .. to 'different con'certs [kɔ̃'n'seʁ] ..oper'as [ɔpe'ʁa] .. and 'shows .. on 'Tuesday 'night in Pa'ris . and we will con'clude with a wonder'ful 'evening in Ver'sailles

(Rogerson-Revell, 2014, p. 148)

Examples 7.1 and 7.2 illustrate that while a level of segmental accuracy is necessary for intelligibility, suprasegmental aspects—including stress, phrasing, speed, and volume—are also essential. Prosodic features not only facilitate speech processing and comprehension, including both lexicogrammatical and pragmatic aspects of utterance meaning, but also enhance the communicative impact of a message (Rogerson-Revell, 2014, p. 154).

One of the benefits of interaction-based research is that such contextualized data illustrates the important role which pronunciation plays not only in intelligibility but also in terms of perceptions of proficiency and competence, and consequently professional image and status (see Rogerson-Revell, 2007). Awareness of this broader impact of pronunciation could help convince business language trainers and their trainees of the value of time spent on this area of language. This point about the broader impact of pronunciation is echoed in research in other areas of professional communication, such as call centers and healthcare.

Call Centers

Increasingly since the 1990s, the call center industry has been outsourced for economic reasons from the United States and the United Kingdom to

countries with lower labor costs, such as the Philippines and India (Friginal, 2007). As major English as a Second Language (ESL) or “outer-circle” (Kachru, 1985) countries with long traditions of bilingual education, these outsourced countries provide a ready source of English-speaking customer service representatives (CSRs) to U.S. and U.K. companies. Typically, this means that CSRs are L2 English speakers, so that the caller and the CSR do not have the same L1 background and are may not be familiar with the accent of their interlocutor.

Issues of intelligibility are a frequent focus in CSR training as are concerns to communicate in a way that will achieve a good level of customer service and avoid frustrating or ineffective service encounters, as illustrated in the following exchange (Example 7.3), where the word *reinstatement* causes communication difficulties between the agent and the customer:

Example 7.3

- Agent: *I'm going to put this through to our reinstatement department*
Customer: *Which depart...which department?*
Agent: *Reinstatement. And they'll be the one to check the record if they could put the policy back in force again.*
Customer: *So it's going to which department now?*
Agent: *Reinstatement department*
Customer: *Statement department?*
Agent: *Reinstatement department*
Customer: *Statement?*
(Agent then spells what he is saying.)

(Forey & Lockwood, 2007, p. 319, emphasis in original)

There is a developing literature in call center research on issues to do with pronunciation (e.g., Hood & Forey, 2008; Lockwood, Forey, & Elias, 2008; Lockwood, Forey, & Price, 2008; Pal & Buzzanell, 2008). Tomalin (2010) investigated communication problems in an offshore call center in India that was doing contract work outsourced from the United Kingdom, finding that a number of pronunciation features caused difficulties—accent (mainly phoneme confusions), stress (e.g., on numbers), pitch and intonation (e.g., use of a “sing-

song” intonation), and speed and pacing (“machine gun” delivery)—in addition to other linguistic features (vocabulary and “Indianisms”) and a factor of culture and empathy that included complex features of voice and prosody. Cowie (2007) and Cowie and Murty (2010) have researched Indian CSRs, focusing on the comprehensibility of varieties of Indian accents for call center work and the extent to which the CSRs accommodated features of the accent of their American callers, such as their pronunciation of postvocalic /r/, into their own speech. Cowie (2007) has also investigated issues relating to identity and accommodation of native speaker accents, claiming that it is common management practice in Indian call centers to seek to disguise national identity through attempts at accent reduction or accent neutralization. Cowie (2007) points out that a new target seems to be emerging for “a ‘neutral’ accent, which in some corporate circles is considered to be a regionless international variety” (p. 316).

Lockwood et al. (2008) analyzed data from an offshore call center in the Philippines, finding that deviation in phonology, both individual sounds and prosodic features and how Philippine English differs from standard varieties of English caused potential misunderstanding between the CSRs and callers, whereas minor grammatical errors did not. Echoing Cowie’s (2007) point, they argue that training for call center representatives needs to focus more on developing a relatively neutral accent that will be understandable to callers from around the world. A study by Yau (2010) of problematic telephone conversations between American customers and Filipino CSRs revealed the importance of callers’ use of prosody and voice quality to convey metessages which the CSRs sometimes were unable to understand or adequately respond to. Yau’s (2010) study suggests that the goal in CSR training must go beyond features of accent and segmental phonology to include both comprehension and productive control of features of prosody and voice that cue interpersonal meaning and interactional dynamics. This presents a necessarily high target for CSR communicative competence that highlights the importance of pronunciation in both its macro and micro dimensions for effective intercultural and international communication (Pennington, 2018b).

In a study of the relationship between language proficiency and success in delivery of telephone support by Filipino CSRs, Friginal (2007) reports that language scores were initially considerably lower than the level of proficiency set by the call center company, although, after training, improvements were made in the CSRs' English during transactions. Two of the four language attributes measured to assess English proficiency relate to pronunciation; these are (1) pronunciation (articulation of segmentals); and (2) application of speech techniques—intonation, pitch and volume, and rate. The other two measured language attributes are (3) vocabulary and grammar; and (4) listening and comprehension.

Friginal (2007, p. 331) points out that language proficiency accounted for only 10% of the variability in the transaction monitoring scores of the participants and that other factors, such as product knowledge, cultural sensitivity, rapport, and personalization of service are important components of effective quality service. At the same time, Friginal suggests that CSRs may benefit from additional training, particularly in the effective use of prosodic aspects of pronunciation. He concludes:

Results suggested an effect of prosody (suprasegmentals) in non-native speech in relation to the service quality scores of CSRs.... It could be inferred that in native speaker and non-native speaker customer service transactions, intonation, pitch and volume, and rate of speech were relevant attributes that raters and customers identified as predictors of success in delivery of service.... This direction for training in the call centers with added focus on prosodic patterns of English might contribute to the development of speaking skills and overall effective delivery of service by Filipino CSRs. (Friginal, 2007, p. 343)

The outsourced call center industry is a massive global enterprise impacting communication across physical, cultural, and linguistic differences. While there are many aspects to developing effective skills for L2 transactions by phone, the significance of pronunciation and its impacts on communication cannot easily be ignored. Rather, this pervasive aspect of communicative performance demands further attention in research, training, and assessment. In the view of Davies (2010), the crucial areas for attention in CSR performance are clarity, intelligibility, and rapport, all of which are impacted by pronunciation.

Health Communication

The internationalization of many workplaces and the linguistic diversity of many populations have heightened awareness of the need to prioritize communication in organizations in order to ensure both successful management and effective service provision. A clear example of this heightened attention to communication is in the provision of healthcare. In many countries, including the United Kingdom and the United States, health services must be provided to a multicultural and multilingual population and increasingly by a linguistically and culturally diverse workforce (Hoekje, 2011).

The field of health communication research has expanded rapidly, as reflected, for example, in the many research centers focused on this area of study in U.K. and U.S. universities. This has created an extensive body of sociolinguistic research into communication in health care settings, particularly, using conversation and discourse analysis approaches (Da Silva & Dennick, 2010; Heritage & Maynard, 2006; O'Grady, 2011), and some employing corpus-based approaches (Adolphs, Brown, Carter, Crawford, & Sahota, 2004; Brown, Crawford, Gilbert, & Gale, 2014). Some studies have looked specifically at language and cultural barriers to effective communication in medical contexts (e.g., Jacobs, Chen, Karliner, Agger-Gupta, & Mutha, 2006; Johnstone & Kanitsaki, 2006), generally focusing on the patients' use of language. John-Baptiste, Naglie, and Tomlinson (2004), for example, found that patients who were less proficient in English were more at risk of physical harm in U.S. hospitals than those who were proficient users of English. Roberts et al.'s (2005) applied linguistic research focused on doctor–patient interactions in 19 London inner-city surgeries and found that 20% of the 232 video recordings were with patients with limited English and contained major and often extended misunderstandings. The researchers concluded that in such episodes, “talk itself is the problem. These misunderstandings related to issues of language and self-presentation rather than culturally specific health beliefs” (Roberts et al., 2005, p. 473). They highlighted four categories of patient talk which caused misunderstandings: (1) pronunciation and word stress; (2) intonation and speech delivery; (3) grammar /vocabulary and lack of contextual information; and (4) culturally specific presentation styles.

To give an example of a pronunciation-specific misunderstanding, the data extract below (Example 7.4) is from a consultation with an Albanian

patient (P) who is trying to explain to the doctor (D) what she thinks is causing a rash. There appears to be confusion caused by the patient's mispronunciation of some vowel sounds in words such as *meat*, *mince*, and *cow*.

Example 7.4 Skin Rash

A young Albanian woman has had a skin rash for several months and thinks it might be an allergy:

- 1 P I think from / meat / because
- 2 D milk
- 3 P / meat /
- 4 D mit what is mit
- 5 P / meat / ((laughs)) I don't know / meat / erm
- 6 D er
- 7 P / mince /¹ I think
- 8 D mice
- 9 P yeah
- 10 D like rat
- 11 P yeah
- 12 D you have mice at home
- 13 P what do you – / mince / no but / meat / you know to eat erm I can't say in English
- 14 D can you draw it
- 15 P no I don't know how to / write / this er
- 16 D is it a food
- 17 P eah food food I think from er from / cow / you know
- 18 D from
- 19 P / cow /
- 20 D coal?
- 21 P no no / cow /
- 22 D cow from cow
- 23 P yes
- 24 D ah beef
- 25 P beef yeah
- 26 D ah beef ah

Roberts et al. (2005) provide further examples of difficulties with intonation and speech delivery. Example 7.5 involves a patient, a Nigerian speaker of English, who was bitten by a dog while on holiday and is concerned about rabies. The researchers maintain that differences in the use of contrastive stress between the patient and the doctor give rise to ambiguity.

Example 7.5 Dog Bite

- 1 D what kind of dog was that (.) it was somebody's (.) [dog]
 2 P [yes] somebody's
 3 D it was a stray dog
 4 P no no it was somebody's dog
 5 D right
 6 P yes I:: made an enquiry they said that they they told me
 7 the dog go to the vet regular
 8 D right okay
 9 P but that's what they said
 10 D right (.) right right so did you know the owner or [did]
 11 P [I] know the owner
 12 D oh fair enough (.) so
 13 P erm:: ((laughs)) (but)
 14 D did you
 15 P no

(Roberts et al., 2005, p. 470)

The patient seems to imply that he is not convinced the dog is free from rabies (lines 6, 7, and 9). However, the stress on *what* in “but that’s what they said” is confusing, since an L1 English speaker would expect the contrastive stress to be on the verb *said*. It appears that the doctor does not pick up the implication of the contrastive stress, with the acceptance “oh fair enough (.) so” (line 12), and the patient tries to show his concern further with the use of a hesitation marker and laughter in line 13 “erm:: ((laughs)) (but).”

Roberts et al. (2005) claim that “the difficulties in struggling to understand such a variety of patient styles and accents was commented upon by GPs [doctors] in video feedbacks, and when we presented to GP groups” (p. 474). They conclude that differences in pronunciation, stress, intonation, and ways of structuring talk can make it difficult for doctors to process information, can lead to long protracted consultations, and can cause considerable uncertainty and frustration.

A similar example (Example 7.6) is reported by O’Grady (2011) in a training roleplay consultation with a (pretend) patient (P) by an (actual) L2 English doctor (D) in a clinical communication workshop for international medical graduates that she observed:

Example 7.6

- D: ... please tell me Margaret have you ever thought of harming yourself
P: No doctor no (.) I’ve never thought about harming myself ((rise-fall tone))
D: Good good but you’re feeling sad (.) for how long have you been feeling sad (p. 58)

O’Grady (2011) comments:

The roleplaying patient’s use of marked stress on *self*, together with a rise-fall tone, marked departure from a simple statement of fact. This prosodic contextualization cue, probably produced without reflection, was a signal to the doctor to explore the patient’s intention and to infer her hidden concern that she might harm her child. It invited an empathic response that would display the doctor’s perception and understanding of the patient’s fear and thus encourage her to elaborate. But here, doctor and “patient” did not share communicative background and did not share the same inferential processes. The doctor did not recognize the prosodic convention and could not infer the patient’s intention. (p. 58–59)

As these two examples illustrate, there is considerable scope for pronunciation-related problems in communication between doctors and patients when one or the other, or both, are L2 speakers.

Historically, a great deal of research into health communication has focused on the patients' language rather than that of healthcare professionals. Awareness is growing however of the need for specific kinds of pronunciation-related communicative competence on the part of medical professionals (Hoekje, 2011; Tipton, 2011), not only in medically sensitive and critical contexts such as giving news of cancer test results (Gillotti, Thompson, & McNeilis, 2002) but also in everyday medical contexts such as those illustrated above where exchanges are between speakers from different linguistic and cultural backgrounds. Doctor-patient exchanges by speakers coming from different linguistic and cultural backgrounds is now common in the United States, which has experienced an acute shortage of medical staff, especially nurses in recent decades, and where "approximately 25% of physicians in training and in practice...are graduates of international medical schools (IMGs)" (van Zanten, 2011, p. 77), and most of these "self-report a native language other than English" (*ibid.*). It is also common in the United Kingdom, where the National Health Service (NHS) draws increasingly on international staff (e.g., doctors, nurses, care workers, hospital support staff) to bolster its workforce. It is estimated, for example, that 25% of doctors, and one in eight nurses, in the United Kingdom do not have English as a first language; yet there seems to be little language support for international staff in the NHS, despite the large number of such employees in the organization (Adolphs et al., 2004).

A key concern in healthcare is to ensure that all staff have adequate language and communication skills to function effectively and safely at work, including those for whom the working language is not their first language. There is considerable evidence that miscommunications between medical staff, other health care personnel, and patients can lead to inappropriate or even harmful treatments (Divi, Koss, Schmaltz, & Loeb, 2007; Flores et al., 2003; Joint Commission on Accreditation of Healthcare Organizations, 2007). In the last decade, the Joint Commission on Accreditation of Healthcare Organizations (2007) annual report on U.S. hospital safety reported that "[i]nadequate communication between care providers or between care providers and patients/families is consistently the main root cause of sentinel events" (p. 46), such as errors in surgery or medication. Graphs in the report (p. 46) show communication as the source of nearly two-thirds

of all of the sentinel events in the period of 1995 to 2006. In a study at a U.S. university medical center, Labov and Hanau (2011) found an alarming number of words dictated by IMGs that were misunderstood by experienced medical secretaries paid to transcribe the dictations, such as *patch* misheard as *plaque*, *ascending* as *descending*, *infarction* as *infection*, and *firm* as *full*. In addition to their pronunciation of phonemes and words impeding intelligibility, IMGs' communication style and English skills affect their clinical performance with, and their evaluation by, patients and medical personnel (Labov & Hanau, 2011, p. 264).

Yet, despite increasing recognition of language-related difficulties in healthcare provision, there has still been little research into the communication and language skills of healthcare professionals, notably international staff. Roberts, Atkins, and Hawthorne (2014) investigated the lower success rate of International Medical Graduates (IMGs) as compared to other candidates in the Clinical Skills Assessment (CSA) part of the examination to become licensed General Practitioners in the United Kingdom.² Their study of performance features in the CSA revealed that candidates' manner, in terms of both voice and body orientation and movement, as well as attitude, were the performance features most noted by examiners. The comments on manner related mainly to how candidates sounded; and although there were no indications that candidates' pronunciation made them hard to understand, many negative judgements having to do with empathy related to features of voice quality or intonation: for example, "same level of voice all the way through—no hint of empathy"; "empathic statements didn't sound empathic." Similar problems in conveying empathy through prosodic features and voice quality have been recognized in research on IMGs in the United States (e.g., O'Grady, 2011). In the U.K. study, the negative perceptions of empathy related to lack of variation in voice level contrasted with positive perceptions of empathy in relation to accent and voice quality, such as "I like the accent (Scottish) sounds empathic," "did soften his voice and show empathy." Roberts et al. (2005) conclude that such "micro-level behavioral cues," which are manifested in how a candidate sounds and which function as highly significant social contextualization cues, can lead to adverse comments about their manner of communicating that can have negative consequences on their skills

assessment. It is this “micro-level,”³ including pronunciation, which Roberts et al. (2014) claim is key to understanding communication issues. They also claim that what they consider these “small differences are the most culturally specific aspects of language” (Roberts et al., 2014, p. 243).

Pronunciation and Image

As we have seen, one of the most important ways people characterize one another is by how they speak, that is, their “style” or “manner” of speaking. People frequently make judgements about others based on initial verbal and visual impressions, such as “there’s something about his tone I didn’t like” or “she’s got a nice manner about her.” As we have seen in Chap. 1, such judgements involve not only linguistic information but also non-verbal cues to make evaluations about aspects of social identity, including class, ethnicity, and occupation, as well as personality characteristics and the communication of emotion and attitude.

The concept of speech style has both an individual and a social component: the speaker can select linguistic elements (grammatical, phonological, and lexical) to show affiliation or identity with a particular social group or to show an individual stance or attitude. In this sense, style is dynamic and creative, the performance aspects reflected in the use of the term **styling** (Coupland, Candlin, & Sarangi, 2001; Rampton, 1999). Styling can be seen as part of the construction of social identity, either by invoking or consolidating membership of a speech community, or by expressing a personal attitude or stance (e.g., distancing oneself) in relation to a specific speech community (Coupland et al., 2001, p. 221).

Vocal Styling

In professional contexts, an effective style or manner of speaking can be not only a powerful tool for getting things done but also a resource for

establishing status and influencing others in developing and managing working relationships or rapport. The ability to style-shift, that is, to adopt, or accommodate to, the communicative and stylistic conventions of a particular interactive context, such as through a shift in voice quality to convey certain types of interpersonal or pragmatic meaning in call center service encounters (Yau, 2010) or through the appropriate use of humor in business meetings (Rogerson-Revell, 2007), is an important professional as well as social skill. Failing to adopt or accommodate to the communicative and stylistic conventions of a particular context, or adopting an inappropriate style, can have a negative effect. For example, Yau (2010) found that the Filipino CSRs she studied were often unable to adopt or accommodate to American customers' shifts in voice quality, with resulting disruption or deterioration in communication.

Cameron (2000) describes the imposition of a particular speech style on a group of U.K. call center workers who are required to talk in a certain way in customer service contexts. Cameron refers to the concept of **vocal styling** to describe how employees are instructed and monitored for their use of politeness phenomena, prosodic, paralinguistic, and non-verbal (body language) behaviors. These include two specific instructions on voice which are invariably given: (a) operators should smile and (b) use "expressive" intonation. Examples of such instructions are "Have a smile in your voice and avoid sounding abrupt" and "...try to make the caller feel you are there for them...avoid a disinterested, monotonous tone of voice" (Cameron, 2000, p. 334). Cameron claims that these and other performance features imposed on the call center workers are typical of a feminine speech style, or what Lakoff (1975) termed "women's language." As Cameron noted, "It has been argued that both smiling and using expressive intonation are symbolically feminine behaviours" (p. 334). Cameron (2000) suggests that such styling in contemporary service workplaces is less a marker of group identity or organizational affiliation than a prescriptive, commercial "branding" practice imposed from the top down.

The role of linguistic styling, including the use of pronunciation and voice quality or delivery, as an image creation technique is not new. It is commonplace, for example, to select a particular voice quality or speech accent in advertisements, to help position and sell a product to a target

audience. For instance, regional or rural accents might be employed to advertise essential, everyday products like butter or bread, while standard, urban accents are employed to advertise aspirational products like perfume or cars. Businesses are now more aware of the fact that certain accents have certain connotations. For example, in the 1990s, the U.K. brewery, Boddington's Bitter, ran an advertising campaign for beer (U.K. "bitter") that challenged preconceptions. The advertisement started by projecting the glamor and mystery of a perfume advertisement, through images, music, and a glamorous model with a breathy, standard GB accent. It turned out she was advertising bitter, switching to a strong Mancunian (from Manchester) accent halfway through the ad. The Mancunian accent is often seen as being industrious or creative while also being warm and welcoming. Research has shown that gender and other aspects of voice quality can be important considerations in people's buying decisions. Wiener and Chartrand (2014) found that female listeners are most likely to purchase a new service when they hear it advertised by a man who has a creaky voice, but least likely to purchase it when they hear it advertised by a woman with a creaky voice. Interestingly, creaky voice, produced with very tense vocal folds, is somewhat the opposite of the qualities of "smiling voice" and expressive intonation which the U.K. call center workers whom Cameron (2000) studied were advised to adopt. Thus, the kinds of voice quality which are effective in business contexts would seem to vary considerably, and blanket prescriptions are risky, given such factors as gender-related differences in conventions and the images which companies try to project for their products and services. These facts have only just started to be investigated.

The relationship between communication and image promotion is also widely acknowledged in the wider public arena, which is why everyone from politicians to pop stars have image advisers, spin doctors, PR consultants, advertising strategists, press spokespersons, media analysts, speech writers, and other marketing specialists on staff. Such specialists know that the way a person speaks in professional contexts affects how the person is judged, especially in terms of persuasion and influence. The "vocal packaging" of individuals, giving them the right personal image, often makes them more alluring and persuasive: vocal styling wins when content alone will not.

Image creation is nowhere more important than in public speaking, where persuasive rhetoric is central to success. Changes in the channels of persuasive talk mean that co-present and live communication with both nearby and mass audiences is generally giving way to distant audiences and delayed communication in broadcast media (Pennington, 2018a). As a result, persuasive talk is becoming more common through mediated dialogue in various forms (such as televised interviews and chat shows) or mediated multi-party talk (such as discussions and press releases) than through live public oratory. Therefore, contemporary public speakers, whether representatives of political parties or of business organizations, need to be able to shift their style of speaking to suit the media situation; they need to be able to deliver a message persuasively in both formal and intimate contexts in order to project the appropriate public or professional image. The role of pronunciation and delivery in achieving this can be illustrated by considering briefly two very public political speakers.

Politically Speaking

Margaret Thatcher, a well-known British political figure, was leader of the right wing Conservative (Tory) party and Prime Minister from 1979 to 1990. Thatcher was widely recognized as an effective and powerful speaker, even by those who opposed her (Atkinson, 1984; Beattie, 1982; Heritage & Greatbatch, 1986; Shea, 1993). However, it is also well documented that Thatcher deliberately “restyled” her voice during her political career in terms of both accent and delivery. Early broadcasts indicate that when she first came to politics she had a normal female, relatively high-pitched voice, which typically got higher when she was emotionally involved or trying to make herself heard. This led her advertising campaign director, Tim Bell, to describe her voice as “schoolmarmish, very slightly bossy, slightly hectoring.... It was a voice from the 1950s that was long gone” (Feldman, 2008). Margaret Thatcher was advised to lower, slow down, and soften her voice to connote greater authority, which she did with vocal training. Over a 10-year period, she managed to lower her pitch a full 60 Hz, which equals half

the difference between the average male voice and the average female voice (Karpf, 2006). She also modified her accent, which initially had some regional characteristics from Lincolnshire in the East Midlands region of England, so that it became increasingly exaggerated RP over the years.

Thatcher also learned to use some rhetorical techniques which are commonly used by politicians and public speakers to create impact. In particular, Thatcher made use of some of the basic rhetorical formats, such as two-part contrasts and three-part lists, that have been found to have a strong, positive influence on audience response (Atkinson, 1984; Heritage & Greatbatch, 1986). As Rogerson-Revell (1995) illustrated, Thatcher displayed her rhetorical skills by combining both two-part contrasts and three-part lists (Example 7.7):

Example 7.7 Margaret Thatcher

- A A and 'had we 'not...
- B and we 'lost ... (2-part contrast)
- 1 either 'Invincible
- 2 or 'Hermes
- 3 or 'both (3-part list)

(Rogerson-Revell, 1995, p. 223)

Thatcher frequently used emphasis and pausing to create an image of authority, using many strong stresses, or prominences, in specific patterns that create “accents of power” (Bolinger, 1986, 1989), together with controlled pausing and shifts in rhythm and key (pitch range), to get and maintain the audience’s attention. Bolinger (1986, 1989) describes how an extra degree of force, or climax, can be achieved in a number of ways that are illustrated in the following passage: by putting a strong stress or accentuating the final word in a phrase (e.g., “had we **not** ... and we **lost**”); by placing strong stresses consecutively, accentuating every word, rather than in alternation (e.g., “**ships do change** direction”); and by adding extra, redundant stresses (e.g., “**criticise criticise criticise**”) (Example 7.8).

Example 7.8 Margaret Thatcher

[*leans forward*] 'ships ↑**do** 'change 'direction\ [*stare*] 'that one had 'zig-zagged 'quite a '**bit**\ .. but you ↑'cannot 'ignore the '**intelligence** \ .. you have ↑'their .. '**big** .. '**cruiser**/ ↑**and** / [*stare*] you 'expected an '**aircraft carrier**/ .. 'soon to be within 'range of [f] ↑'your '**fleet** \ ↓'no I 'think 'these '**criticisms** comes from 'people/ .. who ... ↑'criticise '**criticise** '**criticise**\ .. but 'never had to 'make these '**decisions**\ .[f] [ac] and had we ↑'not\ ... and we '**lost**\ either '**Invincible**/ or '**Hermes**/ or '**both**\ [*stare*] [p] [dc] [*sits back*] ↓ we should have been '**culpable**\ ... 'very ↓'heavily 'culpable\

(Rogerson-Revell, 1995, p. 230)

Table 7.1 Key to transcription conventions (Adapted from Rogerson-Revell, 1995, p. 244)

Prominence	
Bold	Strongly accentuated syllable
'	Stressed syllable
Pitch range	
↓	Marked step-down in key*
↑	Marked step-up in key
* "Key" is used here in a similar sense to Brazil (1994, p. 9) to refer to changes in pitch level.	
Pitch movement (tone)	
\	Final falling tone
/	Final rising tone
∨	Fall-rise
^	Rise-fall
–	Level
Pause	
.	0.5 second pause
..	1 second pause
-er-	Filled pause
<6>	Timed (secs) pause
Pace and volume	
[p]	Soft /low volume
[f]	Loud
[ac]	Accelerated speech
[dc]	Decelerated speech

Rogerson-Revell (1995) provided an example of Thatcher's skill in using delivery to create impact in the section

and had we ↑ **not** \ ... and we **lost** \ either **Invincible**/ or **Hermes**/ or **both** \

where she uses a combination of climatic final stress or accentuation and controlled pauses to create a series of rhythmically similar phrases. Rogerson-Revell (*ibid.*) pointed out that Thatcher also used pauses for dramatic effect in this same passage, with two extra-long post-accentuation pauses, notably the one after *both*, and a noticeable step-up in pitch range or key at the beginning of the section (had we ↑not) and a step-down at the end of it (..or both ↓).

Obviously, audience impressions of speakers vary and, as with Thatcher, can be negative as well as positive. Indeed, Thatcher's vocal styling was not viewed positively by everyone; and some critics have maintained it was so caricatured that it reduced the credibility of her image. As one writer, Keith Waterhouse, explains, "I cannot bring myself to vote for a woman who has been voice trained to speak to me as though my dog has just died" (quoted in Karpf, 2006, p. 34). The emotional response to politicians' speech style is extremely important and, as Waterhouse indicates, can help win or lose elections. Interesting similarities can be drawn between Margaret Thatcher in the United Kingdom and a similarly experienced American politician, Hillary Clinton, the Democratic candidate in the 2016 U.S. presidential elections. Both women adopted the carefully controlled, traditional oratorical style typical of many politicians and both have been criticized on the one hand for their lack of femininity and on the other for shouting, screeching, or sounding too shrill (Chozick, 2016). Clinton's lack of popularity and eventual defeat by Donald Trump has been partially related to their very different speaking styles and public images (Black, 2016).

The maverick style of Donald Trump is reflected both through his accent and his delivery. Regarding his accent, despite being a billionaire, Trump has a working-class New York accent, which he has used to his advantage. Michael Newman, a linguist at Queens College in New York City, is quoted as saying:

Democrat or Republican, in an age where trust in politicians is at a minimum, it is not hard to see the attraction of that blunt aspect of the New York

image.... People do not perceive the New York style of speaking as particularly attractive or high-status. But they do associate it with competence, aggressiveness and directness. (Guo, 2016)

As well as using his working-class accent to show affiliation to working class voters, President Trump has adopted a much more “conversational” style of speaking than most politicians. Rather than the classical rhetorical strategies illustrated in Margaret Thatcher’s or Hillary Clinton’s speech, Trump’s frequently unscripted language is typified by short information units; short, high-frequency words; many intensifiers, such as *tremendous* and *really*; and direct addresses to his interlocutor or audience (e.g. use of *you*), as illustrated in this extract (Example 7.9) from an interview with Donald Trump by Jimmy Kimmel for the American TV program Jimmy Kimmel Live.⁴

Example 7.9 Transcript from Interview with Donald Trump by Jimmy Kimmel

Interviewer: but 'isn't it 'unamerican and '**wrong** ... to 'discriminate 'against '**people** .. 'based on 'their '**religion** ..

Trump: but 'Jimmy ... I'm '**for** it .. but '**look**.... we '**have** .. '**people** .. '**coming**.. 'into our '**country** .. that are '**looking** to 'do .. '**tremendous** .. ↑ '**harm** you 'look at the two .. 'look at ↑ '**Paris** .. 'look at what 'happened in '**Paris** ... *I mean 'these 'people .. they did 'not* come from '**Sweden** ... '**OK** ... 'look at what 'happened in '**Paris** ... 'look at what 'happened 'last '**week** .. in ↑ '**California** ... with with . you 'know ... 'fourteen 'people '**dead** ... 'other 'people '**going** to '**die** ... they're 'so 'badly '**injured** ... we 'have a 'real ↑ '**problem** ..

In this extract, Trump uses certain cues to signal his engagement with his interviewer, for instance, addressing him by name (*Jimmy*), and with the audience, using markers such as *look* and *OK*. Phonologically, the use of short information units, frequent pauses, and stress to add emphasis is noticeable, for example, in the section:

we '**have** .. '**people** .. '**coming** .. 'into our '**country** .. that are '**looking** to 'do .. '**tremendous** .. ↑ '**harm** ...

This emphatic speech style is reinforced by the use of repetition of phrases, such as “look at what happened in Paris” and repetition of words, such as *people*. All of these linguistic and vocal features are characteristic of a **high-involvement conversational style** (see Tannen, 1984) used to create rapport with the audience and show distance from the traditional political establishment. The anti-establishment image which Trump created partly through speech styling, including vocal styling involving his accent and prosody, was central to his political campaign and his electoral success. Pronunciation and vocal cues are only one aspect of the performance factors which can be adjusted to create a particular speech style; but prosodic features in particular, together with other aspects of delivery such as voice quality, speaking rate, and volume, can have a powerful impact in terms of creating a certain image for an audience.

Forensic Linguistics

Another area of professional life where language in general and pronunciation in particular have had considerable impact is **forensic linguistics**. *Forensic* refers to any activity or process that relates to law enforcement or justice systems, so forensic linguistics studies the use of language in a variety of legal processes, from crime scene to law court. Forensic linguistics is a fast growing field within applied linguistics, with its own conferences, journals, and research centers. It is taught as a professional and academic subject in over forty institutions in more than fifteen countries.

Written language is obviously fundamental for how the law is communicated and interpreted, but spoken language can be equally important, such as for analyzing spoken evidence or identifying speakers, by voice, using **ear witnesses** rather than eye witnesses to crimes. Indeed, ear witness evidence using “voice line-ups” has been used in court cases, and this type of naïve speaker identification, in which non-experts listen to a sequence of different voices without being able to see the speakers, has been found to be quite robust (Nolan & Grabe, 1996). It has been argued that oral evidence can be more reliable than written evidence since it is less easy to manipulate or falsify spoken language records than written texts.

However, proving authorship in either medium is not easy. Just as each time people sign their name there will be slight variations in the signature, so there will be small variations in the acoustic message when someone repeats a spoken word or phrase. Nevertheless, **speaker profiling**, involving the analysis of accent or dialect and other aspects of a person's speech patterns, can provide important information about an individual's socio-economic and geographical background, social status, psychological and emotional state, and level of intoxication or stress.

A large amount of forensic linguistic casework is in the area of **forensic phonetics**, such as the analysis of telephone conversations or covert recordings of speech, as this kind of analysis can reveal linguistic cues to authorship, especially through sociolinguistic markers in accents or dialects and speech styles or registers. An example is the notorious case in England of the "Yorkshire Ripper," who killed thirteen women between 1975 and 1980. The crimes were reminiscent of a serial killer in London a century earlier called "Jack the Ripper." A telephone recording was received in 1979 by a man claiming to be "Jack," beginning, "I'm Jack. I see you are still having no luck catching me..." The recording was analyzed by linguist Stanley Ellis, who identified the man's accent as belonging to the town of Sunderland in the northeast of England (French, Harrison, & Windsor Lewis, 2006). The man turned out to be a hoaxer who was eventually arrested in Sunderland in 2007. As a result of the hoax, the real killer was not caught until 2005, as he had initially been eliminated from enquiries on the basis of the dialect evidence (French et al., 2006).

With technical advances in the computer-based analysis of speech, forensic phoneticians have been able to develop their expertise further. In particular, **voice biometrics** (measurements of the characteristics of voices) form the basis of two processes which have a growing range of applications, speaker identification and speaker verification. **Speaker identification** involves identifying an unknown speaker by comparing a voice model or **voiceprint** (spectrogram) of that speaker against a number of other speaker voiceprints. **Speaker verification** or **speaker authentication** compares one voiceprint against another to confirm, or disprove, a match. Both processes are used for forensic purposes, such as in criminal investigations, but speaker verification is also being used increasingly in a gatekeeping function to provide access to secure systems,

such as telephone banking. Many international banking corporations, for example, now use biometric banking software to access online and phone accounts using a specific individual's fingerprint or voice.

Forensic phonetics can also detect a number of behavioral states through voice, such as the voice changes typically associated with psychological stress, including raised pitch, increased vocal intensity, slight increase in speech rate, and disfluencies. One application which is highly prized is the use of speech recognition technology to detect lying. The benefits of this for law enforcement, intelligence services, and related agencies are obvious. A great deal of research has been carried out on deception (DePaulo et al., 2003), notably the analysis of speech and voice to detect lies. Many devices have been developed, and it is claimed that these "truth machines" (ibid.) can be used to detect both psychological stress and lying from the voice. However, independent research carried out on these devices has failed to support such claims. Hollien and Harnsberger (2006) conclude that "there is little evidence to suggest any of these devices are capable of validly detecting psychological stress, truth and/or lying from cues embedded in motor speech" (p. 131).

Speech analysis has also been used for a different type of legal purpose, which is for asylum and immigration screening. This controversial process has been used in several countries, including Canada, Australia, and some western European countries, where consultants are sometimes employed by immigration services to identify the place of origin of refugee or asylum claimants on the basis of their speech, including pronunciation. The process, known as **Language Analysis for the Determination of Origin**, or LADO, is used in cases where further evidence is sought to confirm place of origin by virtue of the claimant's spoken language. Difficulties have arisen however when the system is used to make judgements about nationality, rather than where the person has lived. For example, a court may be trying to decide if a claimant is from Pakistan or Afghanistan: someone from Afghanistan would have a much higher chance of being granted refugee status because of the unrest in that country. Some regional dialects are spoken on both sides of the Pakistan-Afghanistan border, so that making judgements about nationality on the basis of speaking one of these regional dialects would be highly unreliable. This point has been made by forensic linguists, notably, Eades

(2009), who has produced a set of guidelines to be used in such cases and who has pointed out that the mistakes of inexperienced judges can have profound implications for social justice and people's lives. The danger of negative discrimination based on speech or linguistic profiling was illustrated in Chap. 1, for instance, in connection with racial discrimination against African American Vernacular English (AAVE) speakers (Baugh, 2003; see also Labov, 1972).

Speech Technology

The forensic applications of speech and voice analysis have been outlined, but advances in speech recognition and synthesis have wider applications which are having an increasing impact on society and how people live their lives. The field of natural language processing (NLP), involving computer science, Artificial Intelligence (AI), and computational linguistics investigates techniques for recreating or modeling the human processes of speech understanding and speech production using computers.

Speech synthesis, the artificial production of human-like speech (Chap. 5), goes back long before the advent of digital technologies, with early mechanical models using bellows documented in the late eighteenth century. Digital speech synthesis typically involves a text-to-speech (TTS) system which converts normal language text into speech, such as the TTS apps, *KjennMEG* (*KnowME*)⁵ and *Voice4U*,⁶ which enables children with communication difficulties to express themselves. Other systems convert linguistic symbols such as phonetic transcriptions into speech, such as *Phonetizer*.⁷ Speech synthesis has long been an important **assistive technology** tool helping remove communication barriers for people with a wide range of disabilities. One of the earliest applications was screen readers for the visually impaired, but text-to-speech systems are now commonly used by people with dyslexia and other reading difficulties, as well as by pre-literate children. Speech synthesis is also a valuable computational aid for the analysis and assessment of speech disorders and can be used as well to help those with severe speech impairment, by providing a customized, synthesized voice output. A well-known example of a user of this technology was the late Professor Stephen Hawking,⁸ world expert in

cosmology and theoretical physics, who, due to a long-term degenerative disease, lived for many decades largely paralyzed and communicating using a single cheek muscle attached to a speech-generating device. Professor Hawking used speech synthesizers for many years, and despite the opportunity to move to more sophisticated and naturalistic voice synthesizers, he preferred to keep the original synthesized voice that he and many others identified as his own.

Text-to-Speech has many other applications besides assistive technologies. For instance, many educational TTS apps have been developed for educational purposes, such as *Speak It!*⁹ and *Talk to Me*,¹⁰ which aim to help children learn to read and spell. Speech synthesis is also used in second language acquisition. *Voki*,¹¹ for instance, is an educational tool that allows users to create their own talking avatar, using different accents (see Chap. 5). A user's own selected talking avatar can be emailed, embedded on websites, or shared on social media. There are also many web and mobile-based voice synthesizer apps which use TTS to add voice to blogs or webpages, such as *VoiceForge*,¹² and *Type and Talk*,¹³ or apps that change voice features, such as *Voice Changer*,¹⁴ which allows the user to convert text to speech with multiple variations of voice, including gender, age, voice quality, and pitch. The most important qualities of a speech synthesis system are naturalness and intelligibility. Great improvements have been made in both aspects, though the complexities of pronunciation, particularly prosodic aspects, mean that contemporary speech synthesis systems still remain clearly distinguishable from actual human speech.

Many applications now combine speech synthesis and speech recognition (Chap. 5)—for instance, entertainment products such as games, toys, and animations. Speech technology is a major challenge for video game application: recognizing speech from loud, excited, untrained voices of different ages, genders, and language backgrounds in a noisy background. However, voice-activated commands for action and dialogue with virtual characters are offering a new dimension in game play in many gaming genres and across different platforms, including Sony Playstation¹⁵ and Microsoft Xbox.¹⁶

The fastest growing application of speech synthesis and recognition is in **smart technologies**, which are computing systems that use machine learning and complex programming to make decisions without human

intervention. The increasing sophistication of devices, especially smartphones, is already making this a reality and a growing number of people now talk to their mobile smartphones, asking them to send e-mail and text messages, search for directions, or find information on the Web. Sophisticated voice technology is already commonplace in call centers and many companies and public services, such as doctors' surgeries or hospitals, where automated telephone reception systems are used to direct the customer to the required service (often with considerable frustration along the way).

The U.S. company Nuance Communications, which dominates the market for speech recognition, predicts that voice recognition will change the way we communicate with computers. Nuance, which purchased the program, *Dragon Dictate*, developed in the 1990s, and improved and updated it to become *Dragon Naturally Speaking*¹⁷ (Chap. 5), is also one of the architects of Apple's *Siri*,¹⁸ the voice-activated personal assistant that comes built into the latest iPhone. Similar voice functionality is also built into Android, the Windows Phone platform, and most other mobile systems, as well as many apps. With the success of voice activation in mobile phones, companies are now moving their R & D focus towards other areas, particularly televisions and cars. Nuance is currently developing software called *Dragon TV*¹⁹ which conducts voice-activated searches, such as "find movies starring Brad Pitt," without the need for a remote control. A version of this technology is already in some televisions sold by Samsung. Similar technology is already used in voice-activated GPS (or *sat navs*, as they are known in the United Kingdom) and many car manufacturers are building in voice activation functionality into their cars to enable car drivers to find directions, weather information, and music, such as Ford's *Sync*²⁰ software.

One of the difficulties with voice recognition is developing technology that is sophisticated enough to deal with a wide range of languages, accents, and voices. Ford claims on its website that its software recognizes accents across 17 different languages, including regional varieties of German and Chinese, but as far as we know, there is no independent research testing its reliability in the different languages. A considerable amount of research has been carried out on the Nuance *Dragon* system, Apple *Siri*, and similar systems in terms of their value for language-based

tasks, finding that while useful, they have many limitations (for critiques of *Dragon Naturally Speaking* and Apple's *Dictation* software for writing, see Altman, 2013; Hill, 2013).

The complexities of pronunciation and the variety of accents, as reviewed for teaching applications in Chap. 5, pose similar challenges for other types of voice-activated smart applications, such as smart devices in the home, used to control and automate lighting, heating, ventilation, air conditioning systems, and appliances such as washer/dryers, robotic vacuums, or automatic window blinds. Such systems are commonly referred to as the **Internet of Things**, that is, physical devices (e.g., in the home or car) which are connected through smart technologies. An interesting application that has been developed for language learning purposes is the French Kitchen developed by Seedhouse (2017) and colleagues at the University of Newcastle. *The French Digital Kitchen*²¹ tracks actions with motion-sensor technology similar to a Nintendo *Wii*, and uses speech synthesis to help the user prepare a dish. The *Kitchen* was originally developed to enable people with dementia to cook safely in their own homes, but the assistive technology has now been applied to language learning, enabling users to learn French while performing a real-life task, such as cooking a French dish (see also Chap. 5).

Another rapidly developing application of speech synthesis combined with speech recognition is in robotics. An example is the use of robotics as an assistive technology to help deal with the increasing need for nursing support for an expanding ageing population. In the United States, the Nursebot project²² (Pineau, Montemerlo, Pollack, Roy, & Sebastian, 2003) has enabled the development of a mobile robotic assistant to assist elderly individuals with mild cognitive and physical impairments, as well as to provide support for nurses in their daily activities. The researchers found that in an assisted living facility, the robot successfully demonstrated that it could autonomously provide spoken reminders and guidance for elderly residents, such as keeping a doctor's appointment or information about weather reports of television schedules.

Today, robots are designed for specific purposes, as described above, similar to the early development of personal computers. However, with mass production and lowering prices, voice-activated as well as gesture-activated robots will have a substantial impact on society.

Within education, **educational service robots**, which are intelligent robots that have various functions to assist learning, are starting to make an appearance. It is argued by Han (2010) that with their friendly appearance and physical movements, educational service robots can establish interactive relationships with learners, particularly with children, making learning more pleasurable by increasing learners' interest and lowering their "affective filter" (Krashen, 1985) limiting or blocking learning by fear, anxiety, or other forms of negative affect.

Despite the massive and rapid advances in technology, challenges remain with speech synthesis and recognition in terms of the quality of voice input and output. For instance, with voice recognition, there are difficulties dealing with non-standard accents or faulty speech due, for instance, to disabilities or age-related issues. While speech synthesis has improved greatly in recent decades, synthetic voices are generally perceived as hyperarticulated and lacking in expressivity, sounding robotic. One of the continuing challenges therefore is to produce speech synthesizers which can automatically adapt their way of speaking to the contextual and communicative situation, as humans do—at the same time, perhaps, as we humans become more accustomed to mechanical-sounding speech, as millions of people around the world did in the case of Stephen Hawking.

Broader Educational Applications

Educational applications of pronunciation extend beyond second or foreign language teaching and learning to the study and treatment of various kinds of phonological disorders and to the general education of children in L1 contexts. Pronunciation research and practice addresses phonological difficulties faced by children due to developmental delay or other causes and also by adults as a result of trauma or disease. Phonological disorders that impair speech may be diagnosed in school or clinical settings, and are typically remediated through one-on-one instruction and training by a speech therapist. Pronunciation is also relevant to the education of children in L1, as pronunciation underlies, and is interactive with, all aspects of literacy, including vocabulary, spelling, reading, and writing.

Phonological Impairment

Phonological impairment is a cover term for speech disorders involving articulation as well as problems in perceptual discrimination and cognitive processing related to pronunciation. Phonological impairment in childhood can result from many sources (e.g., cerebral palsy, brain infection, or injury) and can also occur later in life, such as a result of hearing loss, stroke, head injury, tumor, or disease or trauma to vocal organs, with varying potential for remediation or improvement. In both children and adults, communication difficulties resulting from phonological impairment can cause sufferers to become anxious, which may make their symptoms worse, and withdrawn, which impacts their relationships with others (Dalton & Hardcastle, 1977; Gibbon, 2007).

Stroke, head injury, or brain disease may result in motor speech dysfunction such as:

- **dyspraxia** or **apraxia**, sometimes referred to as “cluttering” (Dalton & Hardcastle, 1977, chap. 7), caused by a disruption of signals from brain to vocal organs and manifested in excessive speed of speech, which results in slurring, repetition of initial phonemes and syllables, omission and incorrect ordering of phonemes and syllables, and “the telescoping of several syllables of a word” (ibid., p. 109); and
- **dysarthria**, which refers to speech difficulties with a neuromuscular source, manifested in muscle weakness²³ or poor muscle control in speech production, affecting such features as the tension of the tongue muscles required for producing closure and manner distinctions in consonants or of the vocal folds required for voicing distinctions (Dalton & Hardcastle, 1977, p. 30).

In addition, people can lose the ability to read—a condition called **alexia**—as a result of brain trauma or degraded function, such as through injury, stroke, or dementia (National Institute of Neurological Disorders and Stroke, 2017). Another type of phonological impairment is **stuttering** (Dalton & Hardcastle, 1977, chaps. 4 and 6), “a speech disorder characterized by repetition of sounds, syllables, or words; prolongation of

sounds; and interruptions in speech known as blocks,”²⁴ resulting from a complex of motor, cognitive, and psychological factors that generally involve problems in coordination and timing of articulatory plans and their execution (Dalton & Hardcastle, 1977, p. 50ff).

Many types of speech disorder involve disruption in sensory feedback control and motor coordination of articulation (Dalton & Hardcastle, 1977, pp. 30–31). These dysfunctions affect the coordination and order of articulation of sounds and syllables, making the pronunciation of long words especially difficult. Cluttering (dyspraxia or apraxia) and stuttering are sometimes developmental in children, with considerably higher incidence in boys than girls.²⁵ “Stuttering almost always begins in early childhood” (Dalton & Hardcastle, 1977, p. 66) and may or may not be accompanied by cluttering. Cluttering resulting from brain trauma or disease (e.g., Parkinson’s) or psychological disorders is found in both children and adults (Dalton & Hardcastle, 1977, chap. 7).

Even mild hearing loss can affect pronunciation, and studies have also shown that frequent ear infections in childhood can affect pronunciation (Balbani & Montovan, 2003). Partial or complete hearing loss results in delay in the development of language processing and production. In addition, those with hearing loss do not perceive and so typically do not produce voiceless sounds. As another effect, children with hearing loss may not be able to hear themselves speak, and this affects their prosody and voice quality, as they may speak with a voice pitched too high or too low, may speak too loudly or softly, and “may sound like they are mumbling because of poor stress, poor inflection, or poor rate of speaking.”²⁶ All of these areas of pronunciation problems affect deaf people’s self-confidence and ability to communicate and socialize with others, though signing offers a different medium for full communication within the deaf community and among others who have learned to sign.

Phonological impairment in childhood which can involve problems in articulation of consonants (Leonard, 1995) and vowels as well (Gibbon, 2007, p. 250). It may be caused by developmental delay or various organic causes such as cleft palate or deafness but is often of indeterminate origin:

Children with developmental speech disorders form a large, heterogeneous group. Various conditions, both biological and environmental,

place children at risk for speech disorders. These conditions include: sensory deficits (e.g., hearing impairment); cognitive deficits (e.g., learning disabilities, mental retardation); psychiatric and emotional disorders (e.g., autism); neuromotor disorders (e.g., cerebral palsy, Worster-Drought syndrome); and structural abnormalities of the vocal tract (e.g., cleft palate, malocclusion). Developmental phonological disorders...are unlike the disorders listed above because their presence cannot be attributed to any known or detectable condition or cause. (Gibbon, 2007, p. 245)

Developmental phonological disorders, those which have no known cause, are relatively common in childhood (Gibbon, 2007). In many cases, developmental phonological disorders respond to therapy and can be remediated in a matter of months or years, or they may be something a child grows out of by age 5 or 6 (Gibbon, 2007, p. 245). In some cases, however, they present problems beginning in early childhood that persist in the elementary school years and affect a child's self-esteem and ability to communicate and socialize with others. Non-developmental disorders respond to therapy but generally cannot be entirely corrected due to the presence of an underlying deficit or disorder affecting speech. Impairments of phonology can affect children's performance in not just speech but also spelling and reading, and can "have serious implications for future achievement" (Leonard, 1995, pp. 575–576), even in adulthood. In this connection, it can be noted that children who have poor performance on phonological awareness tasks also have reading problems (Menn & Stoel-Gammon, 1995, p. 351), and **dyslexia** is often caused by difficulties in phonological processing (Shaywitz, 1996). Phonological impairment stemming from problems in phonological awareness therefore predicts to dyslexia, though children diagnosed as dyslexic often do not show phonological impairment in their speech. This may be in part because phonological problems are typically noticed at an early age (e.g., at age 3), as a child's pronunciation lags significantly behind other children of the same age, and may be remediated by the time a child is observed to have reading problems more severe than other children of the same age (e.g., at age 7).

The patterns seen in developmental phonological disorder might include, among others, voicing of initial voiceless consonants and stop-

ping of fricatives or liquids, as seen in a child Velleman (1998) calls Nathan (p. 125); “hardening” of final glides or diphthongal elements to stops, as McMahon (2007, p. 172) cites from a study of a child “PS” who pronounced *cow* as [kɒb], *know* as [nəb], and *you* as [jɪb] (p. 172); **metathesis** (interchanging one sound for another), as in [gɒbi] for *buggy*; or **migration** (movement of a sound to a certain position), such as a child who put all initial fricatives in final position, resulting in [nos] for *snow* (Velleman, 1998, p. 123). These patterns can be diagnosed in **phonological process tests** that analyze clients’ phonological patterns in labeling tasks or detailed phonological analysis based on free speech (Velleman, 1998, p. 9). Beyond the kinds of assessments which Velleman (1998) describes, speech therapists and clinicians may observe a client speaking in order to assess articulatory gestures.²⁷

Gibbon (2007) advocates the use of computerized instruments to provide a more accurate picture of clients’ phonological system than can be done by observation or transcription, which “affords at best an indirect representation of the actions of the articulators” (p. 251), whereas “[i]nstrumental procedures, particularly acoustic analysis and electropalatography (EPG), are able to measure objectively aspects of articulation and speech motor control” (p. 252).

EPG is one of the few instrumental techniques able to record directly the actions of one of the major articulators involved in speech production, namely the tongue. EPG records details of the location and timing of tongue contacts with the hard palate during speech. (Gibbon, 2007, p. 252)

Specific technology for this purpose includes the *LinguaGraph*²⁸ system and *LinguaView*²⁹ software using an artificial palate fitted with electrodes (Chap. 5). Gibbon (pp. 252–254) draws attention especially to the phenomena of **covert contrasts** and **undifferentiated gestures**, two aspects of pronunciation in phonological disorder that can be picked up by instruments but not by impressionistic transcription. Covert contrasts are consistently produced distinctions in sounds which speakers are able to make but which are below the level of perception of the transcriber. This is an example in which machine measurement outperforms a human assessor, in its ability to perform a more

accurate micro-level analysis of articulation. Undifferentiated gestures are abnormal articulations that extend beyond specific places of articulation, often spanning large areas of the palate and/or the tongue, and so cannot easily be referenced to traditional places of articulation and transcription symbols. This is an example in which traditional categories such as phonemes and phonetic segments cannot fully capture or accurately describe significant aspects of pronunciation performance.

For some children, only a small portion of their phonological system shows signs of disorder, whereas for others, such as for the child Nathan referred to above, phonological processes affect many different phonemes, and syllable types are highly restricted based on the child's limited phonological repertoire in different positions. An extreme example are the productions by a 5½ year old child Velleman (1998) labels "Jonathan" who pronounced *string* as [wi], *three* as [bi], both *chair* and *crayon* as [de], and both *candle* and *glasses* as [dæ] (Table 3.21, p. 71). The number and types of errors that "Jonathan" makes means that his phonology is not functional, and Velleman sees the goal of the clinician as *Making Phonology Functional*, the title of her book.

The phonological processes and limited set of phoneme contrasts observed in children with phonological disorder, whether they occur in a small or large portion of their phonological system, may be due to mechanical production difficulties such as inadequate control of the tongue (Gibbon, 1998, 2007) for achieving a specific articulatory shape or position or for articulating certain combinations of sounds. However, as Leonard (1995) pointed out, phonological impairment may also involve difficulties of perceptual discrimination, phonological memory, and cognitive processing of linguistic information at various levels, including not only that of syllables and phonemes but also higher word and syntactic levels as these connect to phonology. The sequencing errors of phonological migration and metathesis, for instance, may result from difficulties producing, coordinating, perceiving, and/or remembering articulatory gestures in specific word or syllable positions. As another type of example, difficulties in speech and in spelling and reading may be related to problems in discrimination or awareness of phonemes, as "residual effects" of phonological disorder in the early years that can last

into later childhood, adolescence, and adulthood (Lewis & Freebairn, 1992).

Low phonological awareness is associated with dyslexia in reading as problems of decoding and identifying read words because of difficulty decomposing them into their component phonemes (Shaywitz, 1996). As Koda (2007) points out, “Phonological decoding is a critical component in reading, because it facilitates the extraction and assembly of a word’s phonological information” (p. 222), which is a crucial first step in deciphering the words contained in a written text. Dyslexics’ reading process is labored and can produce incorrect results, as they may retrieve words from memory that have some phonological resemblance to the words they read, such as the same first syllable or overall syllable structure and stress pattern, but are otherwise incorrect. Their reading comprehension process is then quite slow and requires extra time and effort as well to correct wrong word retrievals when these are discovered (e.g., because they do not fit the context further along in the same sentence or text). For this reason, those who suffer from dyslexia may become exhausted and discouraged and so not maintain their reading process to the end of a text.

On the basis of the diagnosis, considering age-appropriate phonological substitutions and patterns versus those which do not occur among other same-age children, the clinician then determines whether a child has disordered phonology and if so, what phonological difficulties to work on. Velleman (1998, p. 10) suggests focusing first on the sounds that are easiest for the child to produce and then on those which are perceptually salient for the child but which the child finds difficult to produce. In her view, this will yield better results than trying to focus the child’s learning on a contrast that is salient for listeners but not (yet) for the child. For adults suffering phonological disorder later in life, diagnosis is more straightforward, assuming that phonology was previously fully mastered, although a thorough screening of individual words and spontaneous speech can help determine exactly what the client’s problems are and which problems to focus on first. In providing therapy for the adult, it makes sense to target therapy first to those sounds that are most salient or problematic from the client’s point of view, that is, those on which the client’s attention is most focused.

For developmental phonological disorder, Gibbon (2007, pp. 263–266) describes four common therapies, which we note can also be applied in L2 teaching:

- *Auditory/perceptual approach*, such as *auditory input therapy*, or *structured listening*, which focuses only on perception, aiming to increase the auditory salience of target speech sounds by presenting them in “contexts that involve maximally clear productions” (p. 264), such as in stressed syllables and focal words.
- *Linguistic/phonological approach*, such as *minimal pair contrast therapy*, using a game format in which a child has to communicate a message to a listener as a way to encourage the child to distinguish words which that child does not ordinarily differentiate. Another approach in this category is that of *maximal/multiple oppositions*, which is based on contrasts not in just one dimension, as is the case with minimal pair contrast therapy, but “along multiple articulatory dimensions of voice, manner, and place of articulation” (p. 264). The pronunciation of maximally differentiated words may in fact be easier for children to master than the pronunciation of the same words taught in a minimal pair.
- *Motor/articulatory approach* is one which “follows general principles of motor learning, which emphasize the importance of providing repetitive, intensive, and systematic practice drills...to establish consistency in articulation and reduce variable performance. Motor approaches emphasize...verbal, visual, tactile, and/or kinaesthetic feedback on performance” (p. 265).
- *Computer-based approaches* using different kinds of material to engage children in perceptual discrimination and identification of phonemes, including synthesized speech making possible “selective cue manipulation” (p. 266) to aid in establishing phoneme boundaries and use of visual EPG feedback to help children visualize tongue gestures and positioning.

Phonics instruction may be of value in phonological disorder as well as in dyslexia and alexia.

For dyspraxia/apraxia (cluttering) and dysarthria, such as after head injury or stroke:

The focus of intervention is on improving the planning, sequencing, and coordination of muscle movements for speech production. The muscles of speech often need to be “retrained” to produce sounds correctly and sequence sounds into words. Exercises are designed to allow the person to repeat sounds over and over and to practice correct mouth movements for sounds. The person with apraxia of speech may need to slow his or her speech rate or work on “pacing” speech so that he or she can produce all necessary sounds. In severe cases, augmentative and alternative communication may be necessary (e.g., the use of simple gestures or more sophisticated electronic equipment).³⁰

Dalton and Hardcastle (1977) note that for cluttering, “simply slowing down has a marked effect” (p. 115), as does syllable-timed speech, “since it discourages the elision and extensive coarticulation of syllables and words, and inevitably slows the rate of speech” (ibid.).

Dalton and Hardcastle (1977) suggest that for most stutterers,

...the emphasis in treatment must be principally on the psychological aspects of fear and avoidance, but they will need some means of modifying their speech behaviour as well as their emotional behaviour.... Briefly, attempts to modify the speech patterns of stutterers fall into three categories: those which seek to replace stuttering by an alternative pattern of speaking, those which aim at reducing the stutter to an easier, more relaxed form, and those which attempt to eliminate disfluencies by inhibition of disruptive elements. (pp 101–102).

As for cluttering, they highlight syllable-timing, in which “the stut-terer is asked to time his speech to a regular and (initially) slow beat on each syllable that he utters” (p. 102), as a main therapy for stuttering. According to the website of the National Institute on Deafness and Other Communication Disorders, therapies for stuttering often focus on timing of speech, speaking slowly, and regulating breathing, in addition to attempting to reduce anxiety, which is related to stuttering as both cause and effect. Drug therapies used to treat anxiety and other disorders are sometimes used as well to treat stuttering, and electrical devices are also used, such as one which “fits into the ear canal, much like a hearing aid, and digitally replays a slightly altered

version of the wearer's voice into the ear so that it sounds as if he or she is speaking in unison with another person. In some people, electronic devices may help improve fluency in a relatively short period of time."³¹

Although phonological impairment has been extensively researched, there is still much that is not known about the many and sometimes complex causes of this condition, and the most effective treatments for the different types of phonological disorder affecting speech production, word recognition and retrieval, and reading comprehension, and for clients of different ages. Brain imaging technologies are being used to explore underlying causes of developmental phonological disorders as well as those phonological disorders resulting from brain injury and disease, and these techniques are being used as well in the study of dyslexia and alexia (National Institute of Neurological Disorders and Stroke, 2017).

One line of ongoing research in developmental phonological disorders, including stuttering, is focused on identifying early signs of the disorders and how they may be treated at the youngest possible age, in order to head off more serious difficulties in communication and social life that affect self-esteem and quality of life, and the effects of phonological disorder on academic and other kinds of achievement. Research is also seeking to determine which children are most likely to overcome phonological disorder or which types of disorder are most treatable, and which need more intensive or long-term therapy. Continuing research is needed in all of these areas, and more research is needed as well on use of instrumentation for assessment and treatment and on the most effective approaches for dealing with children and adult clients and with specific types of phonological disorders. In addition, research horizons for stuttering involve continued testing of electronic devices "to determine how long such effects may last and whether people are able to easily use and benefit from these devices in real-world situations."³² As Gibbon (2007, pp. 266–267) notes, clinicians have become increasingly aware of the need to base their assessments and interventions on research evidence regarding effective methodologies, while also noting a "lack [of] knowledge about the most critical variables for predicting and maximizing progress" (p. 267).

L1 Literacy

As reviewed in Chap. 2, phonological knowledge and processing are related to literacy, and various kinds of interrelationships have been confirmed in research on children. For example, knowledge of phonological patterns as picked up by preschool children in nursery rhymes is reported to be highly predictive of their ability to read by age six, when other characteristics (e.g., IQ, vocabulary, mother's education) are factored out (Goswami & Bryant, 1989). In addition, explicit or controlled phonological processing has been found to be heavily dependent on sound-spelling correspondences learned through reading and reinforced through writing (Reis & Castro-Caldas, 1997). Pronunciation and literacy are moreover connected via auditory sequential memory, which is centrally involved in phonological processing and has also been shown to be related to the spelling abilities of young adolescents (Goyen & Martin, 1977). Indeed, the **phonological loop** (Baddeley, 2012; Baddeley, Gathercole, & Papagno, 1998), which is the mechanism for holding just-heard phonological information in memory and being able to compare it with phonological information in previously stored exemplars, is facilitated by—and may be dependent on—literacy as phonological units and patterns become connected to words.

Vocabulary and Spelling

Phonological learning is based on the child's developing knowledge of words and their sequential patterning (Ellis, 1996, p. 109), and as the size of the child's vocabulary increases, this provides a foundation for developing an increasing knowledge of contextual variants and the abstract phonological categories underlying these (Beckman & Edwards, 2000). Through literacy, the learning of words and the phonological units out of which they are composed become tied to the phonetic values of graphemes and spelling patterns in the written language. In addition, in English, monosyllabic words must be learned in both their full and reduced pronunciations, which often differ substantially (e.g., *a*, *the*, *that*, *and*, *has*, *have*, *do*, *did*), and the differing pronunciations of vowels, con-

sonants, and combinations of these in stressed and unstressed syllables is an aspect of learning new words, both how to recognize them in listening and how to produce them in speaking. An aspect of learning to spell words is learning the pronunciation patterns associated with different vowel and consonant spellings, including combinations of letters with specific pronunciations (e.g., *sh* /ʃ/ as in *shoe*; *ph* /f/ as in *phone*) or variant pronunciations (e.g., *th* /θ/ as in *thin* or /ð/ as in *then*; *x* /ks/ in *box* or /z/ in *xenon*; *gh* /g/ as in *ghost*, /f/ as in *tough*, or silent as in *thought*; *e* /ɛ/ as in *bet* or /i/ as in *be*; *ie* /i/ as in *believe*, *ie* /ɪ/ as in *sieve*, or /aɪ/ as in *tie*). It therefore makes sense when teaching new words to teach their pronunciation in relation to their spelling, and doing so aids reading.

A system for teaching sound-spelling correspondences in English is that of **phonics**, which is widely used as preliminary to teaching reading in the primary school years (grades 1–3) and sometimes also for teaching the sound patterns of English as L2. In contrast to phonetics, a phonics approach starts not from sound but from the written language and the graphic representation of words in relation to their pronunciation. As described by the U.S. National Reading Panel (2000) subgroup reporting on phonics instruction, “Systematic phonics instruction typically involves explicitly teaching students a prespecified set of letter-sound relations and having students read text that provides practice using these relations to decode words” (pp. 2–92). There are various approaches to phonics in use in different countries, such as Australia, the United States, and the United Kingdom. Some emphasize the spelling of individual phonemes whereas others emphasize the spelling of words and parts of words such as the **onset** (initial sound) and **rime** (remainder of the word). What all phonics approaches have in common is that they aim to develop learners’ phonological awareness in relation to their lexical knowledge and their ability to read written language.

According to the report by the U.S. National Reading Panel (2000), based on a meta-analytical review of a substantial body of research carried out largely in the 1990s, “systematic phonics instruction makes a bigger contribution to children’s growth in reading than alternative programs providing unsystematic or no phonics instruction” (pp. 2–92), and “phonics instruction produces the biggest impact on growth in reading when it begins in kindergarten or 1st grade before children have learned to read

independently” (pp. 2–93). These positive findings were confirmed in later reviews of research carried out in the United Kingdom, including Torgerson, Brooks, and Hall (2006) and Rose (2006). A review of reading programs for children published on the Best Evidence Encyclopedia website of Johns Hopkins University suggests that while phonics is part of effective teaching approaches, the specific instructional approaches used are an important factor in the effectiveness of reading programs (Slavin, Lake, Chambers, Cheung, & Davis, 2009).

Reading

Learning to read in one’s native language starts from learning correspondences between the pronunciation of known words and their spelling, then gradually acquiring knowledge of the sound-spelling correspondences of the language that are related to previously learned articulatory patterns and phonemic categories of the spoken language. The knowledge of sound-spelling correspondences gained from known words forms a basis on which the learner can build a larger vocabulary of words met first in reading. For those words known first in reading, their pronunciation may only later (if at all) be encountered in spoken language contexts. Although the pronunciation of a word can be guessed at based on common sound-spelling correspondences, the actual (conventional or “correct”) pronunciation may not follow the common pattern and so it often happens that people mispronounce words they encounter only in reading.

As many studies have demonstrated, reading is heavily dependent on phonological awareness (e.g., Adams, 1990; Ehri, 1994, 1998; Koda, 2007; Menn & Stoel-Gammon, 1995; Shaywitz, 1996). According to Koda (2007):

Phonology plays a pivotal role in a learner’s process of establishing systematic linkages between spoken language elements and graphic symbols. Phonology continues to be essential in reading and the processing of printed text, well beyond the initial stages in the acquisition of literacy, because visually presented information must be converted into its phonological form in order to be stored and processed efficiently in working memory. Since virtually all of the sub-component processes of comprehension rely

on working memory, phonological processing remains critical in text understanding at all stages of reading development. Thus, phonology is essential in acquiring literacy, supporting and promoting it in many ways. (p. 219)

Koda cites research showing that

- children's sensitivity to the structure of spoken sounds is directly related to their ability to read and spell words;
- phonological segmentation capability is a powerful predictor of reading success among early and middle-grade students; and
- reading progress is significantly enhanced by phonological awareness training. (pp. 220–221)

The inter-influence of phonology and reading goes both ways, as experience reading builds increasing knowledge of word structure and sound structure in language (Koda, 2007, p. 222).

According to Koda (2007), “in all languages, phonological decoding is an indispensable competence for reading acquisition and comprehension” (p. 228), though the sound-to-grapheme correspondences on which phonological decoding is based vary from one language to another. For this reason, phonological decoding skills and processing routines developed for a specific L1 do not automatically apply to another language, whose orthographic system may differ in ways both subtle and substantial, as differences in pronunciation for a particular grapheme across languages can be slight or major. In addition, some L1 graphemes—and indeed, whole orthographic systems—will have no equivalent in the learner's L2 and so the potential for applying phonological decoding skills and processing routines from the L1 to the L2 is reduced, such as for Chinese (as L1 or L2) in relation to any other language that does not use Chinese characters, or *hanji*. Yet, as a general principle, learners will transfer what they know from L1 to the processing of L2, thus transferring their phonological and related orthographic knowledge of their L1 to their L2 performance. When the systems of graphic symbols for representing sounds are similar in the L1 and the L2, as they are for those languages using the Roman alphabet, this reinforces the potential for

transfer, as the processing of L2 can be carried out to a significant extent using L1 decoding skills and phonological processing routines.

Writing

The effect of pronunciation on writing can be seen in errors that derive from pronunciation, so called “pronunciation-spelling” errors (see Treiman, 1993, for examples and discussion of these in children’s spelling), such as substituting *of* for *have* in the perfect modals *could have*, *should have*, *would have*; intuitive sound-based spelling (e.g., “meledy” for *melody*, “jeddy” for *jetty*); dialectal pronunciation-based spelling (e.g., “perdict” for *predict*, “aks” or “axe” for *ask*; “pritty” for *pretty*); or L2 spelling based on the influence of L1 pronunciation and spelling. These may be **cognitive errors** (Sterling, 1983) stemming from the fact that the child does not know how the word is spelled and simply tries to work it out based on pronunciation or stemming from cognitive activity while writing that mistakenly retrieves a spelling pattern based on spoken language. The occurrence of such sound-based errors in students’ written work might justify attention to pronunciation in relation to English orthographic patterns as feedback on their writing.

Concluding Remarks

The discussion and examples in this chapter, like other parts of this book, illustrate the central role of language in people’s lives and particularly the crucial significance of pronunciation in a wide range of social, professional, and educational contexts. Among these contexts, second language learning and teaching is obviously key but there are many other domains where spoken language, and pronunciation in particular, plays a crucial role. All communication, including spoken communication, is an interactive process and effective interaction is a joint enterprise between participants. Success depends on a variety of factors, including cultural background, linguistic skill and repertoire, personal experiences, attitudes, and expectations. There can be considerable variation in these

factors across individuals, events, and contexts, with consequent variability in terms of communicative success.

It is intriguing to consider whether societal and technological developments will lead to changes in how and when spoken language is used. We are already seeing the growing use of machine-based speech and speech recognition in various aspects of our daily lives. Similarly, changes in the modes and channels of mass communication, particularly the development of social media tools such as YouTube and vlogging (video blogging) may result in the evolution of speech and pronunciation styles to suit new media and new purposes. Whatever changes in communication are coming, it seems unlikely that speech, and pronunciation specifically, will become any less central than it is now; and arguably oral communication could become even more important than it is today—perhaps as a needed balance or a correction to the heavy reliance in the present era on short-form written communication such as texting. What is clear is that there will be a continuing need to understand and be understood by a wide variety of people in a wide range of contexts, including the many subtleties of understanding people's social selves and the metamessages they communicate to others through pronunciation.

As the world becomes smaller and societies and workplaces become more diverse, the importance of raising awareness of effective communication increases. People need to be able to accommodate linguistic and phonological variation, and they need to know how to speak as clearly and effectively as possible to a potentially highly linguistically diverse audience. This returns to the issue raised multiple times in this book regarding the desirability of a common phonological core that allows for significant accent variation in ELF versus a relatively “neutral” EIL accent that is widely known and shared in order to ensure intelligibility across speaker groups and cultures. The possibility and the value of a “universal” EIL accent, as advocated by many of those working in CALL center pronunciation research, must be balanced against the reality of language change and people's deep-seated need for linguistic distinctiveness, as they express their identity through communicative style and accent, even in relatively restricted contexts such as call center business transactions and customer service interactions. Furthermore, as research has shown,

there is not a one-to-one correlation between accent and intelligibility, and being able to speak clearly does not necessarily mean eroding one's accent. Greater awareness of such issues might help combat the broader stigma of foreign-accented speech and reduce some of the broader social consequences of negative attitudes.

Notes

1. In British English, *mince* refers to ground beef.
2. The Step 2 Clinical Skills (CS) performance test is also part of the United States Medical Licensing Examination “required of all physicians who seek graduate training positions in the United States, including graduates from medical schools located in the United States and other countries” (van Zanten, 2011, p. 78). The Step 2 CS exam evaluates candidates on the dimensions of Spoken English Proficiency, Communication and Interpersonal Skills, and Integrated Clinical Encounter (see Chap. 6).
3. We note that Roberts et al.'s (2014) conception of difficulties in pronunciation that include prosody and voice quality as “micro-level” communication and language skills is quite a different conception from our division of pronunciation competence into a “micro” level comprising skills of pronunciation mechanics and discrimination that are required for intelligibility, and a “macro” level that incorporates pragmatics, social meaning, and interpersonal dynamics (see, e.g., the discussion in Chap. 6). Indeed, it can be maintained that the sorts of features of prosody and voice that Roberts et al. (2014) have identified are overarching, and in this sense “macro” features of communication that affect the interpretation of all other (i.e., lexicogrammatical) features, making it possible to correctly interpret them and generally being more impactful than the specific words or grammar employed.
4. Jimmy Kimmel Live TV show (Dec 8, 2015) <https://www.youtube.com/user/JimmyKimmelLive/videos>
5. <https://www.facebook.com/kjennmeg>
6. <https://voice4uac.com/>
7. <https://www.phonetizer.com/ui>
8. <http://www.hawking.org.uk/>
9. <http://www.speakit.info/>

10. <http://www.rosettastone.co.uk/>
11. <https://www.voki.com/>
12. <https://www.voiceforge.com/>
13. https://play.google.com/store/apps/details?id=com.parth.type_talk&hl=en
14. <https://play.google.com/store/apps/details?id=com.androidrocker.voicechanger&hl=en>
15. <https://www.playstation.com/en-gb/>
16. <https://www.xbox.com/en-gb>
17. <https://www.nuance.com/en-gb/dragon.html>
18. <https://www.apple.com/uk/ios/siri/>
19. <https://www.nuance.com/en-gb/mobile/mobile-solutions.html>
20. <https://www.ford.com/technology/sync/>
21. https://europeandigitalkitchen.com/?page_id=402
22. <http://www.cs.cmu.edu/~flo/scope.html>
23. <https://www.asha.org/public/speech/disorders/ApraxiaAdults/#tx>
24. <https://www.nidcd.nih.gov/health/stuttering>
25. <https://www.nidcd.nih.gov/health/apraxia-speech>
26. <https://www.asha.org/public/hearing/Effects-of-Hearing-Loss-on-Development/>
27. <https://www.nidcd.nih.gov/health/apraxia-speech>
28. <https://www.eda.kent.ac.uk/medical/linguagraph.aspx>
29. <https://www.eda.kent.ac.uk/medical/linguaview.aspx>
30. <https://www.asha.org/public/speech/disorders/ApraxiaAdults/#tx>
31. <https://www.nidcd.nih.gov/health/stuttering>
32. <https://www.nidcd.nih.gov/health/stuttering>

References

- Adams, M. J. (1990). *Beginning to read*. Cambridge, MA: MIT Press.
- Adolphs, S., Brown, B., Carter, R., Crawford, C., & Sahota, O. (2004). Applying corpus linguistics in a health care context. *Journal of Applied Linguistics*, 1(1), 9–28. <https://doi.org/10.1558/japl.1.1.9.55871>
- Altman, J. (2013). Taming the dragon: Effective use of dragon naturally speaking speech recognition software as an avenue to universal access. *Writing & Pedagogy*, 5(2), 333–348. <https://doi.org/10.1558/wap.v5i2.333>
- Atkinson, J. M. (1984). *Our masters' voices*. London: Routledge.

- Baddeley, A. D. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, 63, 1–29. <https://doi.org/10.1146/annurev-psych-120710-100422>
- Baddeley, A. D., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, 105(1), 158–173. <https://doi.org/10.1.1.464.9511&rep=rep1>. Also available at <https://www.ncbi.nlm.nih.gov/pubmed/9450375>
- Balbani, A. P. S., & Montovan, J. C. (2003). Impact of otitis media on language acquisition in children. *Jornal de Pediatria*, 79(5), 391–396. <https://doi.org/10.1590/S0021-75572003000500005>. English version retrieved January 10, 2018, from <http://www.scielo.br/pdf/jped/v79n5/v79n5a05.pdf>.
- Baugh, J. (2003). Linguistic profiling. In S. Makoni, G. Smitherman, A. F. Ball, & A. K. Spears (Eds.), *Black linguistics: Language, society and politics in Africa and the Americas* (pp. 155–168). Oxon, UK and New York: Cambridge University Press.
- Beattie, G. W. (1982). Turn-taking and interruption in political interviews: Margaret Thatcher and Jim Callaghan compared and contrasted. *Semiotica*, 39(1/2), 35–45. <https://doi.org/10.1515/semi.1982.39.1-2.93>
- Beckman, M. E., & Edwards, J. (2000). The ontogeny of phonological categories and the primacy of lexical learning in linguistic development. *Child Development*, 71(1), 240–249. <https://doi.org/10.1111/1467-8624.00139>
- Black, E. (2016, June 23). Comparing the speaking styles of Clinton and Trump – And what it reveals about their minds. *Minnesota Post*. Retrieved from <http://www.minnpost.com/eric-black-ink/2016/06/comparing-speaking-styles-clinton-and-trump-and-what-it-reveals-about-their-minds>
- Bolinger, D. (1986). *Intonation and its parts: Melody in spoken English*. Stanford: Stanford University Press.
- Bolinger, D. (1989). *Intonation and its uses: Melody in grammar and discourse*. Stanford: Stanford University Press.
- Brazil, D. (1994). *Pronunciation for advanced learners of English*. Cambridge: Cambridge University Press.
- British Council. (2013). The English effect. Retrieved from <http://www.british-council.org/organisation/policy-insight-research/research/the-english-effect>
- Brown, B. J., Crawford, P., Gilbert, P., Gilbert, J., & Gale, C. (2014). Practical compassions: Repertoires of practice and compassion talk in acute mental

- healthcare. *Sociology of Health and Illness*, 36(3), 383–399. <https://doi.org/10.1111/1467-9566.12065>
- Brutt-Griffler, J. (2008). Intellectual culture and cultural imperialism: Implications of the growing dominance of English in academia. In C. Gnutzmann (Ed.), *English in academia: Catalyst or barrier?* (pp. 59–72). Tübingen: Gunter Narr.
- Cameron, D. (2000). *Styling the worker: Gender and the commodification of language in the globalised service economy*. *Journal of Sociolinguistics*, 4(3), 323–347. <https://doi.org/10.1111/1467-9481.00119>
- Chozick, A. (2016, February 4). Hillary Clinton raises her voice, and a debate over speech and sexism rages. *The New York Times*. Retrieved from <http://www.nytimes.com/2016/02/05/us/politics/hillary-clinton-speeches-sexism.html>
- Coupland, N., Sarangi, S., & Candlin, C. (Eds.). (2001). *Sociolinguistics and social theory*. London: Longman.
- Cowie, C. (2007). The accents of outsourcing: The meanings of “neutral” in the Indian call center industry. *World Englishes*, 26(3), 316–330. <https://doi.org/10.1111/j.1467-971X.2007.00511.x>
- Cowie, C., & Murty, L. (2010). Researching and understanding accent shifts in Indian call centre agents. In G. Forey & J. Lockwood (Eds.), *Globalization, communication and the workplace: Talking across the world* (pp. 125–144). London: Continuum.
- Da Silva, A. L., & Dennick, R. (2010). Corpus analysis of problem-based learning transcripts: An exploratory study. *Medical Education*, 44(3), 280–288. <https://doi.org/10.1111/j.1365-2923.2009.03575.x>
- Dalton, P., & Hardcastle, W. J. (1977). *Disorders of fluency and their effects on communication*. London: Edward Arnold.
- Davies, A. (2010). Language assessment in call centres: The case of the customer service representative. In G. Forey & J. Lockwood (Eds.), *Globalization, communication and the workplace* (pp. 242–248). London: Continuum.
- DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, 129(1), 74–118. <https://doi.org/10.1037/0033-2909.129.1.74>
- DeShields, O. W., Jr., & de los Santos, G. (2000). Salesperson’s accent as a globalization issue. *Thunderbird International Business Review*, 42(1), 29–46. [https://doi.org/10.1002/1520-6874\(200001\)42:1<29::AID-TIE3>3.0.CO;2-P](https://doi.org/10.1002/1520-6874(200001)42:1<29::AID-TIE3>3.0.CO;2-P)
- Divi, G., Koss, R. G., Schmaltz, S. P., & Loeb, J. M. (2007). Language proficiency and adverse events in U.S. hospitals: A pilot study. *International*

Journal for Quality in Health Care, 19(2), 60–67. <https://doi.org/10.1093/intqhc/mzl069>

- Eades, D. (2009). Testing the claims of asylum seekers: The role of language analysis. *Language Assessment Quarterly*, 6(1), 30–40. <https://doi.org/10.1080/15434300802606523>
- Ehri, L. C. (1994). Development of the ability to read words: Update. In R. Ruddell, M. Ruddell, & H. Singer (Eds.), *Theoretical models and processes of reading* (4th ed., pp. 323–358). Hillsdale, NJ: Lawrence Erlbaum.
- Ehri, L. C. (1998). Grapheme-phoneme knowledge is essential to learning to read words in English. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 3–40). Mahwah, NJ: Lawrence Erlbaum.
- Ellis, N. C. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition*, 18(1), 91–126. <https://doi.org/10.1017/S0272263100014698>
- Feldman, S. (2008, September 5). Speak up. Why do women screech when men shout? *New Humanist*. Retrieved September 22, 2017, from <https://newhumanist.org.uk/articles/1858/speak-up>
- Flores, G., Barton Laws, M., Mayo, S., Zuckerman, B., Abreu, M., Medina, L., et al. (2003). Errors in medical interpretation and their potential consequences in pediatric encounters. *Pediatrics*, 111(1), 6–14. <https://doi.org/10.1016/j.annemergmed.2012.01.025>
- Forey, G., & Lockwood, J. (2007). “I’d love to put someone in jail for this”: An initial investigation of English needs in the business processing outsourcing (BPO) industry. *English for Specific Purposes*, 26(3), 308–326. <https://doi.org/10.1016/j.esp.2006.09.005>
- Forey, G., & Lockwood, J. (Eds.). (2010). *Globalization, communication and the workplace: Talking across the world*. London: Continuum.
- French, J. P., Harrison, P., & Windsor Lewis, J. (2006). Case report. R-v-humble, J. S. The Yorkshire Ripper Hoaxer trial. *International Journal of Speech, Language and the Law*, 13(2), 255–273. <https://doi.org/10.1558/ijssl.2006.13.2.255>
- Friginal, E. (2007). Outsourced call centers and English in the Philippines. *World Englishes*, 26(3), 331–345. <https://doi.org/10.1111/j.1467-971X.2007.00512.x>
- Gibbon, F. E. (1998). *Lingual articulation in children with developmental speech disorders*. PhD thesis, University of Luton (now Bedfordshire).
- Gibbon, F. E. (2007). Research and practice in developmental phonological disorders. In M. C. Pennington (Ed.), *Phonology in context* (pp. 245–273). Basingstoke: Palgrave Macmillan.

- Gillotti, C., Thompson, T., & McNeilis, K. (2002). Communicative competence in the delivery of bad news. *Social Science & Medicine*, 54(7), 1011–1023. [https://doi.org/10.1016/S0277-9536\(01\)00073-9](https://doi.org/10.1016/S0277-9536(01)00073-9)
- Goswami, V., & Bryant, P. (1989). *Phonological skills and learning to read*. Hove, UK: Lawrence Erlbaum.
- Goyen, J. D., & Martin, M. (1977). The relation of spelling errors to cognitive variables and word type. *British Journal of Educational Psychology*, 47(3), 268–273. <https://doi.org/10.1111/j.2044-8279.1977.tb02355.x>
- Guo, E. (2016, February 9). Donald Trump's accent, explained. *The Washington Post*. Retrieved July 2, 2017, from <http://www.washingtonpost.com/news/wonk/wp/2016/02/09/whats-up-with-donald-trumps-voice/>
- Han, J. (2010). Robot-aided learning and r-learning services. In D. Chugo (Ed.), *Human-Robot Interaction*. Retrieved August 15, 2017, from http://cdn.intechopen.com/pdfs/8632/InTech-Robot_aided_learning_and_r_learning_services.pdf
- Heritage, J., & Greatbatch, D. (1986). Generating applause: A study of rhetoric and response in party political conferences. *American Journal of Sociology*, 92(1), 110–157. <https://doi.org/10.1086/228465>
- Heritage, J., & Maynard, D. (Eds.). (2006). *Communication in medical care*. Cambridge, MA: Cambridge University Press.
- Hill, C. J. (2013). Apple's dictation software: A voice solution for writers whose hands need a rest. *Writing & Pedagogy*, 5(2), 346–355. <https://doi.org/10.1558/wap.v5i2.349>
- Hoekje, B. J. (2011). International medical graduates in U. S. Higher education: An overview of issues for ESP and applied linguistics. In B. J. Hoekje & S. M. Tipton (Eds.), *English language and the medical profession: Instructing and assessing the communication skills of international physicians* (pp. 3–19). Bingley, UK; Leiden and Boston: Emerald; Brill.
- Hoekje, B. J., & Tipton, S. M. (Eds.). (2011). *English language and the medical profession: Instructing and assessing the communication skills of international physicians*. Bingley, UK; Leiden and Boston: Emerald; Brill.
- Hoffman, S., & Mai, R. (2014). *Accents in business communication: An integrative model and propositions for future research*. *Journal of Consumer Psychology*, 24(1), 137–158. <https://doi.org/10.1016/j.jcps.2013.09.004>
- Hollien, H., & Harnsberger, J. D. (2006). Voice stress analyzer instrumentation evaluation. *Final report for Department of Defense counterintelligence field activity contract*—FA 4814-04-0011. <https://doi.org/10.1.1.525.3830&rep=rep1>

- Hood, S., & Forey, G. (2008). The interpersonal dynamics of call centre interactions: Constructing the rise and fall of emotion. *Discourse and Communication*, 2(4), 389–409. <https://doi.org/10.1177/1750481308095937>
- Jacobs, E., Chen, A. H., Karliner, L. S., Agger-Gupta, N., & Mutha, S. (2006). The need for more research on language barriers in health care: A proposed research agenda. *The Milbank Quarterly*, 84(1), 111–133. <https://doi.org/10.1111/j.1468-0009.2006.00440.x>
- Jenkins, J. (2000). *The phonology of English as an international language*. Oxford: Oxford University Press.
- Jenkins, J. (2014). *English as a lingua franca in the international university: The politics of academic English language policy*. Abingdon, UK: Routledge.
- John-Baptiste, A., Naglie, G., & Tomlinson, G. (2004). The effect of English language proficiency on length of stay and in-hospital mortality. *Journal of General Internal Medicine*, 19(3), 221–228. <https://doi.org/10.1111/j.1525-1497.2004.21205.x>
- Johnstone, M. J., & Kanitsaki, O. (2006). Culture, language, and patient safety: Making the link. *International Journal of Quality Health Care*, 18(5), 383–388. <https://doi.org/10.1093/intqhc/mzl039>
- Joint Commission on Accreditation of Healthcare Organizations. (2007). *Improving America's hospitals: The Joint Commission's annual report on quality and safety*. Retrieved January 10, 2018, from http://www.jointcommission.org/improving_americas_hospitals_the_joint_commissions_annual_report_on_quality_and_safety_-_2007/
- Kachru, B. B. (1985). Standards, codification and sociolinguistic realism: The English language in the outer circle. In R. Quirk & H. Widdowson (Eds.), *English in the world: Teaching and learning the language and literatures* (pp. 11–30). Cambridge: Cambridge University Press.
- Karpf, A. (2006). *The human voice (how this extraordinary instrument reveals essential clues about who we are)*. New York: Bloomsbury.
- Koda, K. (2007). Phonology in literacy. In M. C. Pennington (Ed.), *Phonology in context* (pp. 219–244). Basingstoke, UK and New York: Palgrave Macmillan.
- Krashen, S. (1985). *The input hypothesis: Issue and implications*. New York.
- Labov, W. (1972). *Language in the inner city: Studies in the Black English Vernacular*. Philadelphia: University of Pennsylvania Press.
- Labov, J., & Hanau, C. (2011). Pronunciation as life and death: Improving the communication skills of non-native English speaking pathologists. In B. J. Hoekje & S. M. Tipton (Eds.), *English language and the medical profession:*

- Instructing and assessing the communication skills of international physicians* (pp. 261–285). Bingley, UK; Leiden and Boston: Emerald; Brill.
- Lakoff, R. (1975). *Language and women's place*. New York: Harper and Row.
- Lalwani, A. K., Lewin, M., & Li, K. L. (2005). Consumer responses to English accent variations in advertising. *Journal of Global Marketing*, 18(3/4), 36–47. https://doi.org/10.1300/J042v18n03_07
- Leonard, L. B. (1995). Phonological impairment. In P. Fletcher & B. MacWhinney (Eds.), *The handbook of child language* (pp. 573–602). Oxford, UK and Cambridge, MA: Blackwell.
- Lewis, B., & Freebairn, L. (1992). Residual effects of preschool phonology disorders in grade school, adolescence, and adulthood. *Journal of Speech and Hearing Research*, 35(4), 819–831. <https://doi.org/10.1044/jshr.3504.819>
- Lockwood, J., Forey, G., & Elias, N. (2008). Call center measurement processes in non-English speaking contexts. In D. Belcher (Ed.), *English for specific purposes in theory and practice* (pp. 143–164). Ann Arbor: University of Michigan Press.
- Lockwood, J., Forey, G., & Price, H. (2008). Englishes in the Philippine business processing outsourcing industry: Issues, opportunities and research. In M. L. S. Bautista & K. Bolton (Eds.), *Philippine English: Linguistic and literary perspectives* (pp. 157–172). Hong Kong: Hong Kong University Press.
- McMahon, A. (2007). Sounds, brain, and evolution: Or why phonology is plural. In M. C. Pennington (Ed.), *Phonology in context* (pp. 159–185). Basingstoke: Palgrave Macmillan.
- Marsh D., & J. Laitinen (2005). Medium of instruction in European higher education: *Summary of research outcomes of European Network for Language Learning amongst Undergraduates (ENLU) Task Group 4*. Jyväskylä, UniCOM, University of Jyväskylä.
- Menn, L., & Stoel-Gammon, C. (1995). Phonological development. In P. Fletcher & B. MacWhinney (Eds.), *The handbook of child language* (pp. 335–359). Oxford, UK and Cambridge, MA: Blackwell.
- National Institute of Neurological Disorders and Stroke. (2017). Dyslexia Information Page National Institutes of Health. Retrieved August 17, 2017, from <https://www.ninds.nih.gov/Disorders/All-Disorders/Dyslexia-Information-Page>
- National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Report of subgroups, Chap. 2, Alphabeticity – Part 2: Phonics instruction (pp. 2–88–2–176). Rockville, MD: National Institute of Child Health and Human Development. Retrieved August 10, 2017, from <https://www1.nichd.nih.gov/publications/pubs/nrp/Documents/report.pdf>

- Nolan, F., & Grabe, E. (1996). Preparing a voice line-up. *International Journal of Speech, Language and the Law*, 3(1), 74–94. <https://doi.org/10.1558/ijssl.v3i1.74>
- O’Grady, C. (2011). Teaching the communication of empathy in patient-centered medicine. In B. J. Hoekje & S. M. Tipton (Eds.), *English language and the medical profession: Instructing and assessing the communication skills of international physicians* (pp. 43–71). Bingley, UK; Leiden and Boston: Emerald; Brill.
- Pal, M., & Buzzanell, P. M. (2008). The Indian call center experience: A case study in changing discourses of identity, identification, and career in a global context. *Journal of Business Communication*, 45(1), 31–60. <https://doi.org/10.1177/0021943607309348>
- Pennington, M. C. (2018a). *Connecting everywhere and nowhere: How mass media and digital technology are affecting communication*. Unpublished ms.
- Pennington, M. C. (2018b). Pronunciation and international employability. In *AESLA (Spanish Association of Applied Linguistics) Conference XXVI*, April 2018, Cadiz, Spain.
- Phillipson, R. (2008). *Lingua franca or lingua frankensteinia?* English in European integration and globalisation. *World Englishes*, 27(2), 250–267. <https://doi.org/10.1111/j.1467-971X.2008.00555.x>
- Pineau, J., Montemerlo, M., Pollack, M., Roy, R., & Sebastian, T. (2003). Towards robotic assistants in nursing homes: Challenges and results. *Robotics and Autonomous Systems*, 42(3–4), 271–281. [https://doi.org/10.1016/S0921-8890\(02\)00381-0](https://doi.org/10.1016/S0921-8890(02)00381-0)
- Rampton, B. (1999). Styling the other: Introduction. Special issue on Styling the Other, *Journal of Sociolinguistics*, 3(4), 421–427. <https://doi.org/10.1111/1467-9481.00088>
- Reinares-Lara, E., Martín-Santana, J. D., & Muela-Molina, C. (2016). The effects of accent, differentiation, and stigmatization on spokesperson credibility in radio advertising. *Journal of Global Marketing*, 29(1), 15–28. <https://doi.org/10.1080/08911762.2015.1119919>
- Reis, A., & Castro-Caldas, A. (1997). Illiteracy: A cause for biased cognitive development. *Journal of the International Neuropsychological Society*, 3(5), 444–450. Retrieved August 15, 2017, from <https://www.cambridge.org/core/journals/journal-of-the-international-neuropsychological-society/article/illiteracy-a-cause-for-biased-cognitive-development/880830EAC2B2A5778A33C7F450927A34>

- Roberts, C., Atkins, S., & Hawthorne, K. (2014). Performance features in clinical skills assessment. Centre for Language, Discourse & Communication, King's College London. Retrieved August 10, 2017, from <http://www.kcl.ac.uk/sspp/departments/education/research/Research-Centres/ldc/publications/index.aspx>
- Roberts, C., Moss, B., Wass, V., Sarangi, S., & Jones, R. (2005). Misunderstandings: A qualitative study of primary care consultations in multilingual settings, and educational implications. *Medical Education*, 39(5), 465–475. <https://doi.org/10.1111/j.1365-2929.2005.02121.x>
- Rogerson-Revell, P. (1995). Speaking with style: Delivery and speaker impact in public discourse. In P. Bruthiaux, T. Boswood, & B. Du-Babcock (Eds.), *Explorations in English for professional communication* (pp. 215–244). Hong Kong: City University of Hong Kong, Department of English.
- Rogerson-Revell, P. (2007). Humour in business: A double-edged sword. A study of humour and style shifting in intercultural business meetings. *Journal of Pragmatics*, 39(1), 4–28. <https://doi.org/10.1016/j.pragma.2006.09.005>
- Rogerson-Revell, P. (2008). Participation and performance in international business meetings. *English for Specific Purposes*, 27(3), 338–360. <https://doi.org/10.1016/j.esp.2008.02.003>
- Rogerson-Revell, P. (2010). Can you spell that for us non-native speakers? Accommodation strategies in international business meetings. *Journal of Business Communication*, 47(4), 432–454.
- Rogerson-Revell, P. (2014). Pronunciation matters: Using English for international business communication. In R. van den Doel & L. Rupp (Eds.), *Pronunciation matters: Accents of English in the Netherlands and elsewhere* (pp. 137–159). Amsterdam: VU Uitgeverij.
- Rose, J. (2006). *Independent review of the teaching of early reading*. Final Report, March 2006. Washington, DC: Department for Education and Skills. Retrieved July 20, 2018, from <http://dera.ioe.ac.uk/5551/2/report.pdf>
- Seedhouse, P. (Ed.). (2017). *Task-based language learning in a real world digital environment: The European digital kitchen*. London: Bloomsbury.
- Seidlhofer, B. (2003). A concept of International English and related issues: From 'real English' to 'realistic English'? *Language Policy Division DG IV—Directorate of School, Out-of-School and Higher Education Council of Europe*, Strasbourg. Retrieved from <http://www.coe.int/T/DG4/Linguistic/Source/SeidlhoferEN.pdf>
- Shaywitz, S. E. (1996, November). Dyslexia. *Psychology Today*, 98–104. Retrieved January 10, 2018, from http://dyslexia.yale.edu/PAR_DysArticleintro.html

- Shea, M. (1993). *Personal impact: The art of good communication*. London: Mandarin.
- Slavin, R. E., Lake, C., Chambers, B., Cheung, A., & Davis, S. (2009, June). Effective beginning reading programs: A best-evidence synthesis. Best Evidence Website. Johns Hopkins University School of Education, Center for Data-Driven Reform in Education. Retrieved February 22, 2018, from http://www.bestevidence.org/word/begin_read_Jun_23_2009.pdf
- Sterling, C. (1983). Spelling errors in context. *British Journal of Psychology*, 74(3), 353–364. <https://doi.org/10.1111/j.2044-8295.1983.tb01867.x>
- Tannen, D. (1984). *Conversational style: Analyzing talk among friends*. Oxford: Oxford University Press.
- Tipton, S. M. (2011). The clinical skills (CS) test: IMG preparation and perception. In B. J. Hoekje & S. M. Tipton (Eds.), *English language and the medical profession: Instructing and assessing the communication skills of international physicians* (pp. 91–110). Bingley, UK; Leiden and Boston: Emerald; Brill.
- Tomalin, B. (2010). India rising: The need for two way training. In G. Forey & J. Lockwood (Eds.), *Globalization, communication and the workplace* (pp. 172–189). London: Continuum.
- Torgerson, C. J., Brooks, G., & Hall, J. (2006). A systematic review of the research literature on the use of phonics in the teaching of reading and spelling. *DfES Research Report, No. 711*. London: Department for Education and Skills.
- Treiman, R. (1993). *Beginning to spell: A study of first-grade children*. New York and Oxford: Oxford University Press.
- van Zanten, M. (2011). Evaluating the spoken English proficiency of international medical graduates for certification and licensure in the United States. In B. J. Hoekje & S. M. Tipton (Eds.), *English language and the medical profession: Instructing and assessing the communication skills of international physicians* (pp. 75–90). Bingley, UK; Leiden and Boston: Emerald; Brill.
- Velleman, S. L. (1998). *Making phonology functional: What do I do first?* Boston: Butterworth-Heinemann.
- Wiener, H. D., & Chartrand, T. (2014). The effect of voice quality on ad efficacy. *Psychology and Marketing*, 31(7), 509–517. <https://doi.org/10.1002/mar.20712>
- Yau, J. N. W. (2010). Call centre discourse: Graduation in relation to voice quality and attitudinal profile. In G. Forey & J. Lockwood (Eds.), *Globalization, communication and the workplace* (pp. 107–124). London: Continuum.



8

Relating Pronunciation Research and Practice

Introduction

As we have stressed throughout this book, pronunciation is central to communication and impacts many aspects of daily life, not only in the sense of how people understand each other verbally but also in the image and sense of identity conveyed through voice and accent. We have also made the point that linkages between research and practice are formed in a two-way process in which each feeds the other and contributes to building a knowledge base and theory. Researchers have a role and responsibility to provide rigorous data to help pronunciation practitioners make informed decisions about how and what to teach. Practitioners, on the other hand, need to be able to take a step back from their own teaching to consider the research base supporting what they do in their classes and also to systematically evaluate their own methodologies and practices and how these are impacting their students. For this purpose, pronunciation teachers and practitioners are well placed to conduct classroom-based or field-based action research. There is also an important intermediary role for those who work between teaching and research contexts, with a foot in each of those worlds.

We saw in Chaps. 3 and 4 how research evidence can help in deciding appropriate pronunciation models and goals as well as in selecting teaching priorities, approaches, and specific classroom methodologies. In Chap. 5, we examined the wide range of technologies that can be applied in pronunciation teaching and the research supporting their use. In Chap. 6, we considered the complex area of pronunciation assessment and how research is taking this field forward. In Chap. 7, we reviewed various areas of research that are helping to identify aspects of pronunciation needing attention in broader workplace contexts. In this chapter, we review what we consider to be some of the most significant areas of current research and highlight the gaps that remain between pronunciation research and practice. We also consider some common debates or controversies which have arisen because of these gaps. Finally, we outline what we see as important and promising areas for future research while also considering ways to ensure that pronunciation research and practice remain in a continually productive relationship.

Key Research Themes

There is now a substantial and growing body of research into L2 speech, sociolinguistics, and communication in a range of contexts which offers valuable insights about pronunciation. Although not directly applicable to or focused on pronunciation teaching and learning, research that incorporates data analysis and theorizing on the functioning of pronunciation in real-life contexts provides an important backdrop of knowledge informing pronunciation practices in the language classroom and wider world as reviewed in the chapters of this book. In addition, there is also a substantial and growing body of research and practice specifically focused on pronunciation. As noted earlier, the field of applied phonology, particularly researching the teaching and learning of pronunciation, is relatively new and just starting to crystallize into some specific investigative orientations and directions, such as the research on pronunciation in contexts of ELF, international business and other ESP contexts such as healthcare, and the orientation of many studies on FFI teaching initiatives.

At a general level, it is possible to relate research in pronunciation pedagogy and practice to three broad questions: what to teach, why to teach, and how to teach effectively. This last question of how to teach effectively relates to another area of enquiry that impacts on all of the other questions, which is what are teachers' and learner's conceptions, beliefs, and attitudes about pronunciation (see Table 8.1).

Regarding the question of what to teach, most teachers and teaching materials assume the need to prioritize pronunciation content. Yet questions of what to teach are far from resolved. The issue arises as to what model for pronunciation is appropriate, both in the sense of what variety or varieties to teach and in the sense of which aspects of pronunciation content to prioritize. We saw in Chap. 3 that research in several areas, including intelligibility, functional load, ELF, and the pragmatic and social functions of pronunciation, have a direct relevance to deciding teaching priorities for pronunciation. However, there are as yet only limited links between research in these areas and instructional practice and materials in the teaching of pronunciation.

A similar point regarding the lack of linkage between research and practice can be raised concerning the question of why teach pronunciation, particularly relating to learner needs and achievable, appropriate

Table 8.1 Pedagogical research themes

Questions	Focus
What aspects of pronunciation to teach	<i>Content</i> —Functional load, ELF, intelligibility, segmental/suprasegmental aspects, pragmatic and social functions
Why teach pronunciation	<i>Aims & Purposes</i> —Learner needs/wants, achievable goals, appropriate models
How to teach pronunciation effectively	<i>Approaches</i> —Effectiveness studies, top-down and bottom-up approaches, FFI, proprioception, CAPT and individualization, autonomous or integrated approach
What are teachers' and learners' conceptions, beliefs, and attitudes towards pronunciation	<i>Psychological Context</i> —Knowledge of pronunciation features and their impact (lexical, grammatical, and communicative), myths and misconceptions, role of accent, identity, teaching/learning style, willingness to communicate

goals and outcomes. There appears to be a mismatch to some extent between teachers' and students' recognition of the importance of pronunciation in language learning and both the amount of time invested in pronunciation syllabuses and assessments and the focus of pronunciation instruction. While research shows that there is not a complete interdependence between intelligibility and accentedness, teaching, assessment, and research often assumes a close correspondence between the two concepts. In addition, aspects of pronunciation that go beyond intelligibility per se to affect communicative dynamics and audience impact more broadly are often not recognized as part of what needs to be taught.

The questions of what aspects of pronunciation to teach and how to teach pronunciation are far from settled and depend on many contextual factors. Research has only scratched the surface in determining which orientations and methodologies are most effective for different learner needs and circumstances. Clearly, further work needs to be done regarding what and how to teach pronunciation, but arguably the areas that require the most attention from researchers relate to the effectiveness of pronunciation teaching and to teachers' and learners' knowledge, beliefs, and attitudes regarding pronunciation. As Levis (2016) suggests, "a lot of energy has been spent on discussing what to teach and not enough attention is given to *how* (e.g. is technique X more effective than technique Y: what do teachers actually do when they teach pronunciation) and *who* (e.g. why do some students improve while other similar students do not)" (p. 6). This points up the need for research specifically focused on pronunciation teaching effectiveness.

In addition to these matters of pedagogy, there are a number of active areas of applied research involving pronunciation assessment that require further attention, involving the questions of what is assessed, how it is assessed, and how effectively assessment is carried out in both human and machine ratings of pronunciation for L2 learning and assessment of pronunciation in clinical settings and workplace environments. Basic questions of the nature of pronunciation also impact language testing theory and practice, as they impact theories of language learning.

New lines of research in psychology and related fields on language aptitude, memory, and cognitive processing referenced in Chap. 2 intersect applied work in pronunciation in many ways that have implications for language learning, teaching, and testing. Studies examining personality,

identity, and social orientation in relation to pronunciation performance, aptitude, memory, and cognitive processing are creating new frontiers for research that are shedding light on the nature of pronunciation in relation to individual differences. Such studies are so far few in number, and the relationships between cognitive and psychological variables, on the one hand, and language learning behaviors and outcomes, on the other, are not yet well established. Moreover, the implications for instruction of findings relating to individual differences in pronunciation performance have hardly been explored. Hence, the findings of these studies are as yet only suggestive and indicative of the need for more research to understand the cognitive and psychological bases of pronunciation and how these might be addressed through instruction.

Gaps Between Pronunciation Research and Practice

Many applied linguists aim to apply their research primarily, although not exclusively, to language teaching. However, the link between research and the practical concerns of language teachers, curriculum developers, assessors, and materials writers is not always clear. In the area of pronunciation teaching, the gap between research and practice seems more evident than in other areas, such as grammar and vocabulary (Derwing & Munro, 2015; Levis, 2005). It also seems to be an area of language that creates considerable concern, unease, and debate among teachers (Jenkins, 2012; Rogerson-Revell, 2011). Levis (2005) suggests that the disconnect between research and teaching means that teachers have to intuitively decide which pronunciation features have the greatest effect on clarity and which are learnable in a classroom setting. We would add that teachers' lack of knowledge about the role of pronunciation in wider aspects of communication that go beyond clarity, aspects that have been extensively researched in conversational and workplace contexts, is an important limiting factor on instruction.

It is suggested that the lack of synergy between research and pedagogy inhibits awareness of significant findings which can influence decision making at the levels of pronunciation syllabus design, teaching, and testing. It is

also suggested that this gap between research and practice encourages the perpetuation of approaches to teaching and testing which may not always be appropriate in an era when English is increasingly used by speakers of other languages for international communication. In the following sections, we will consider some of the most significant areas of pronunciation where there appears to be a gap between research and practice.

Research into L2 Speech and Pronunciation Teaching

Since the latter part of the twentieth century, a significant research community has developed focusing on adult phonetic learning, such as the acquisition of specific English phonemes (e.g., of /r/ and /l/ by Japanese learners; Saito, 2011; Saito & Lyster, 2012; Strange & Dittmann, 1984) and research into age of L2 learning (AOL) and foreign accent (Flege, Munro, & MacKay, 1995; Saito, 2015). Much of this research has involved laboratory-based studies conducted by psycholinguists, acoustic phoneticians, or cognitive linguists that have been published in theoretically or technically oriented speech journals. Although many of these studies are relevant and useful for second language teaching and learning, most have escaped the attention of applied linguists and pronunciation practitioners. Fortunately, the more recent studies are starting to appear in SLA and language learning journals. However, these are not necessarily the journals that are read by practitioners.

Practitioners' lack of access or attention to research journals has resulted in a research–practice gap, particularly in terms of awareness by teachers of the cognitive and developmental aspects of phonological acquisition. An example is the importance of phonetic perception in L2 pronunciation learning. Iverson and Kuhl's (1995) “perceptual magnet effect” (Chap. 2) claims that the L1 sound categories form prototypes and, in the words of Leather (1999), these “L1 prototypes constrain learners' abilities to perceive contrasts in L2 by the ‘pull’ they exert” (p. 5). This concept relates directly to the critical listening approach proposed by such practitioners as Couper (2011, 2015), which seeks to adjust category boundaries between phonemes and figure–ground relationships in perception as established in a learner's L1. It could also be related to the pedagogical orientation proposed

by Underhill (2012), a pronunciation teacher educator who stresses the need to “liberate the learner from the *oral* and *aural* grip of their mother tongue pronunciation” by focusing on the 500-year-old notion of “*proprioception*, ...our internal kinesthetic awareness of the position and movement of our muscles and parts of the body” (emphasis in original). It would add authority to Underhill’s advocacy of teaching based on notions of proprioception, which was reviewed in Chap. 4, to relate theory (e.g., regarding proprioception or related concepts such as the perceptual magnet effect) more explicitly to practice, as Couper (2015) has attempted to do to support his recommendations for use of critical listening (though not as far as we are aware in relation to the perceptual magnet effect).

A different kind of example is the concept of functional load developed by theoretical linguists of the Prague School in the early part of the twentieth century as a motivating factor in sound change (e.g., Jakobson, 1931) and later seen as relevant to pronunciation acquisition by some applied linguists and pronunciation practitioners (e.g., Brown, 1991; Catford, 1987; Munro & Derwing, 2006; Rogerson-Revell, 2011). However, there is still little evidence of the concept being used to inform pronunciation teaching and materials or effectiveness research. In this case, as in others, a main reason is the lack of connection between the concerns of linguistic researchers and those of teachers and learners.

Pronunciation Teaching Practices

While there is a substantial body of research into L2 speech and pronunciation specifically, much of it is not directly related to pronunciation teaching. Research into actual teaching practices for pronunciation was until the last few decades quite limited. Researchers are now taking more interest in teaching practices, including how pronunciation is taught, what knowledge and skills teachers have, and how effective pronunciation instruction is for improving intelligibility and other aspects of communicative effectiveness. This interest in pronunciation teaching by researchers that is creating linkages between research and practice is at the same time building theory–practice connections, as researchers attempt to ground their methodologies and explanations for their findings in theories of language learning and teaching.

Teacher Surveys

There has been a variety of surveys of English pronunciation teaching; some in L1 English countries such as Canada (Foote, Holtby, & Derwing, 2011) and Ireland (Murphy, 2011) and some in non- L1 English countries such as Spain (Walker, 1999) and Poland (Gonet, Szpyra-Kozłowska, & Świąciński, 2010). Perhaps one of the most comprehensive and recent is the “English Pronunciation Teaching in Europe Survey” (Henderson et al., 2012, 2015). This large-scale online survey of language teachers in seven European countries revealed some interesting comparisons and generalizations regarding approaches to pronunciation teaching. The researchers found that in the seven countries surveyed (Finland, France, Germany, Macedonia, Poland, Spain, and Switzerland), there was little evidence of practical pronunciation training for teachers, with many teachers relying on theoretical knowledge of phonetics and phonology gained from their university study (Henderson et al., 2012, 2015). For instance, one Spanish teacher stated: “We had a few classes about the pronunciation of English, intonation etc. but just the theory and no actual demonstration of how to teach them” (Henderson et al., 2012, p. 14). The researchers also claimed that among teachers “there is also a widespread opinion that having good pronunciation is sufficient for teaching pronunciation” (ibid.).

In both Henderson et al.’s (2012, 2015) European survey and in a similar survey of teachers in Ireland (Murphy, 2011), researchers found a propensity to focus on traditional teaching methods such as reading aloud and getting learners to mimic the teacher’s pronunciation, and what Murphy refers to as “a noticeable lack of innovation and diversity in pronunciation teaching” (p. 13). Henderson and Jarosz’s (2014) study of how English pronunciation is covered in school textbooks in France and Poland revealed that the majority of the exercises focused on segmentals and provided very little in the way of communicative practice.

Many of these surveys indicate that teachers often feel under-prepared or ill-equipped to teach pronunciation. They also suggest that many teachers tend to adopt a rather limited range of pronunciation task types, such as reading aloud, listen and repeat, or minimal pair practice, based on traditional classroom textbooks (Derwing, Diepenbroek, & Foote,

2012; Tergujeff, 2013). Such findings seem to indicate a lack of awareness of current research and recent approaches or resources for teaching pronunciation, or a tendency to disregard research findings and to stick with older approaches, on the part of many teachers of English. This reflects the fact that many teachers may not have confidence or sufficient grounding in research to go in new directions and the related issue that many teachers have little opportunity for professional development in the area of pronunciation teaching.

Classroom-Based Research

Much of the research into teaching and learning practices has been survey- or interview-based, with relatively few classroom or observation-based studies. This is not surprising given the difficulties of conducting classroom-based research, such as gaining ethical approval, finding and recruiting an appropriate sample of learners, and carrying out experimental and longitudinal studies. Nevertheless, such studies are essential in order to provide evidence of actual teaching practices and learning performance.

It is refreshing to note a number of classroom-based pronunciation studies, especially instructional innovation and intervention studies with an experimental or quasi-experimental design taking place in regular, intact classrooms (e.g., Gordon, Darcy, & Ewert, 2013; Sato & Lyster, 2012), as well as studies in which students are recruited to participate in teaching initiatives offered in laboratory-classroom settings (e.g., Saito, 2011, 2013; Saito & Lyster, 2012). Such studies, together with the findings of investigations employing methodologies such as observation of teaching practices or collection of data by questionnaires, interviews, or focus groups which supplement and complement the information obtained from studying instructional interventions, offer important insights into which approaches to pronunciation teaching are and might be effective in regular classroom instruction, thus offering guidance for teaching practice.

One recent observation-based study by Foote, Trofimovich, Collins, and Urzúa (2013) reported on 400 video-recorded hours of classroom language teaching of three experienced teachers in Québec, Canada. Results showed that pronunciation teaching accounted for only 10%

of total teaching time and focused almost exclusively on spontaneous correction of individual phonemes, with no teaching of suprasegmentals. The authors also claim that the findings of their observations do not confirm findings from teacher surveys regarding the quantity and quality of pronunciation teaching, leading them to conclude “that teachers are not focusing on pronunciation as much as they think they are” (Foote et al., 2013, p. 191). This misalignment between teachers’ and learners’ perceptions of pronunciation practice and actual classroom practice suggests a gap in current research, as there are very few studies that have documented L2 pronunciation teaching practices through classroom observation. Future studies could therefore beneficially combine interview or survey data of teachers’ and learners’ beliefs and practices with analyses of their actual classroom behavior.

Teachers’ and Learners’ Conceptions and Attitudes

As with other areas of language, a growing body of research in pronunciation is emerging which recognizes the influence of social and psychological factors in learning and achievement (reviewed in Chap. 2). As we have seen before, pronunciation is closely bound up with identity and tends to elicit stronger affective responses than other linguistic levels such as grammar and vocabulary, including stereotyping and prejudice (see Chap. 1). In addition, both teachers’ and learners’ conceptions of pronunciation are often restricted views centered on decontextualized notions of phonemic accuracy or correctness based on a native speaker model and often incorporating incorrect information, such as about a necessary linkage of rising and falling sentence-final tones to grammatical meaning.

There have been many studies of learners’ attitudes and beliefs about pronunciation teaching and learning, for instance, focusing on issues of identity and attitudes to accent (Cutler, 2014; Derwing, 2003; Rindal, 2013) and motivation (Moyer, 2007). Attitudes towards pronunciation play an important role in determining learning choices and outcomes. In many cases, learner perceptions are corroborated across studies, for instance, regarding learners’ generally positive attitude towards pronunciation as a focus of instruction and their desire for more pronunciation practice than is actually provided (Foote et al., 2011). However, there are

instances where findings from studies of learners' attitudes and beliefs about pronunciation have produced contrasting results. For example, several studies have reported that learners aspire towards a nativelike (typically, an educated British or North American) English accent (e.g., Derwing, 2003; Janicka, Kul, & Weckwerth, 2005; Nowacka, 2012). In a study of university students of English in Italy, Spain and Poland, for example, Nowacka (2012) claims that most of the students wanted to sound "native-like":

Questionnaire and experimental research clearly shows that to most learners, at least in the European context, correct native(-like) pronunciation is not only a question of communicative pragmatics, but also self-image. And listeners, both native and non-native, evaluate the speaker on the basis of his [*sic*] pronunciation. (p. 56)

However, this preference for a native-like pronunciation goal for European learners is not reflected in some other studies. For instance, in Rindal's (2010) study of Norwegian adolescent learners of English, "a large minority report a wish to avoid native accents and use a *neutral* variety of English" (p. 211, emphasis in original). Tergujeff's (2013) study of Finnish learners shows a similar sentiment, as one respondent states in response to a question about the desire to have a nativelike accent:

Ei se hienoa olis. Haluan korostaa sitä että en ole brittiläinen vaan olen suomalainen.

("No, it wouldn't be nice. I want to emphasise that I'm not British but a Finn.")

(Valtteri, lower secondary level) (p. 84)

This ambivalence to native speaker goals is reflected in Rogerson-Revell's (2010) research into L2 speakers' attitudes to pronunciation when using English for international business communication. In such contexts, it appears that L2 business professionals typically have a pragmatic approach to pronunciation and while some aspire to a nativelike pronunciation goal, many are motivated by "communicative efficiency" and are tolerant of accents as long as they do not impede intelligibility. This is consistent with the English as a Lingua Franca (ELF) orientation of work by Jenkins (2000) and others (e.g., Cruttenden, 2014; Seidlhofer,

2011). As a contrasting case, some employers of L2 speakers handling overseas phone calls from both L1 and L2 English-speaking clients and some researchers studying communication in these contexts (Chap. 7) are calling for a more “neutral” or “accent-free” form of English as needed in these contexts (e.g., Cowie, 2007; Lockwood, Forey, & Price, 2008).

These different contexts and their sometimes conflicting requirements for communication suggest the need for a broader research focus. This would involve investigating not only traditional L2 learners of English but also users of English who may communicate regularly, and effectively, in English but are not highly motivated to improve their pronunciation, as well as those who may be motivated by the communication requirements of their work to aim for expert-level pronunciation competence. In this connection, it has been argued that the increasing importance of the role of English as a lingua franca calls into question traditional notions of motivation and in particular the concept of integrative motivation, as discussed in Chap. 2. Dörnyei and Csizér (2002) argue that the growth of World English “undermines the traditional definition of integrativeness as it is not clear any more who the ‘L2 speakers’ or the members of the L2 community are” (p. 453). Recent research has suggested that a useful way of viewing the complexities of motivation is through the “willingness to communicate” (WTC) framework initiated by McCroskey and Richmond (1987) and later applied to language learning by MacIntyre, Clement, Dörnyei, and Noels (1998). The approach foregrounds the significance of personal attributes and variables, such as personality, confidence, and competence in motivating an individual to speak to another (see Chap. 2). To date, limited research has been conducted in this area in pronunciation teaching and learning, although Derwing, Munro, and Thomson (2008) found some differences in WTC between two immigrant groups in Canada, one Chinese and the other Slavic. Their findings suggest that greater WTC helped the Slavic group make more contact with interlocutors and more progress in their pronunciation.

Further research is needed into this and other aspects of pronunciation, particularly investigating the attitudes and motivations of teachers and learners, taking account of a wide range of sociocultural variables. As was concluded from the review of research on individual differences in L2 learning in Chap. 2, the contribution of sociocultural factors in pronun-

ciation learning is an area ripe for further study. In addition, the relative contributions of nature and nurture, that is, of individual characteristics such as aptitude and personality and how these interact with learning opportunities and environments to create differences in pronunciation outcomes, have hardly been explored. Like most forms of instruction, the emphasis in pronunciation teaching has traditionally been on the class as a whole rather than on individual learner characteristics. However, research findings suggesting the importance of individual motivations and characteristics in pronunciation outcomes challenges the effectiveness of many kinds of whole-class instruction and supports individualized learning environments such as are provided in online computer-based teaching hardware and software (Chap. 5). This is yet another area in which more synergy between research findings and teaching environments would be beneficial.

Pronunciation Textbooks and Other Course Materials

Pronunciation textbooks and other course materials often show gaps between research and practice and between theory and practice. Based on the need to have a wide audience appeal, published pedagogical materials in all fields, such as textbooks and online courses, tend to be conservative and to reflect established teaching traditions and understandings of a discipline that are sometimes outdated and no longer considered by experts to be correct. In addition, if written by practitioners who do not have a strong empirical orientation or research connection, they may promote teaching methods that are ungrounded with respect to accepted theory and research findings. Technological resources can suffer from the same shortcomings and also from limitations of the technology, as well as from some designers and technicians' interest in and limited knowledge of education and the subject which they are designing and programming applications for. Textbooks and other materials for teaching pronunciation are subject to these limitations, as we have seen for technology in Chap. 5.

There is some evidence that the gap between research/theory and teaching is starting to narrow in the area of published teaching materials. For example, several well-regarded applied linguistics books on phonology,

such as Roach's (2009) *English Phonetics and Phonology: A Practical Course*, Collins and Mees' (2013) *Practical Phonetics and Phonology*, and Cruttenden's (2014) *Gimson's Pronunciation of English*, have included sections specifically on language teaching and learning. Several other books directly apply theoretical aspects of phonetics and phonology to pronunciation teaching (e.g., Brown, 2014; Derwing & Munro, 2015; Rogerson-Revell, 2011), and some pronunciation technologies also show similar connections to theory—for example, Cauldwell's (2012) *Cool Speech* app, which is based in an empirically and theoretically grounded model of the communicative functions of pronunciation. These books and technologies can be considered to provide “best practice” exemplars which, it is hoped, will offer guidance for continuing to codify theoretical and research knowledge within pronunciation teaching textbooks and digital resources.

The Effectiveness of Pronunciation Teaching

As with other areas of language learning, an important question for pronunciation practitioners is how effective pronunciation instruction is. Obviously, one of the main difficulties with any effectiveness study is that it requires some measurement of improvement in performance, whether in receptive or productive skills or both. Many earlier effectiveness studies were laboratory-based, such as de Bot and Mailfert's (1982) and de Bot's (1983) research into intonation teaching, although, as we stated earlier in this chapter, there has been a gradual shift in focus over recent decades towards more classroom-based research.

Perceptual training is one area of pronunciation teaching which has long been seen as central to enabling learners to hear and distinguish new sounds, as an essential receptive dimension of pronunciation competence. We saw in Chap. 3 that a particular type of perceptual training, high variability perceptual/phonetic training (HVPT), appears to be particularly effective and can also facilitate the development of productive skills (Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999; Logan, Lively, & Pisoni, 1991; Thomson, 2011; Wong, 2012). HVPT exposes learners to multiple speakers and variations of target sounds rather than limiting input to the productions of a single

speaker and one consistent model for target sounds. Thomson's (2011) study of Mandarin speakers' acquisition of ten English vowels, for instance, provides some evidence of the effectiveness of HVPT for improving vowel intelligibility. HVPT is one area of computer-assisted pronunciation training (CAPT) with considerable potential for a beneficial effect on learning; however, there is limited information to date to confirm or support the effectiveness of CAPT. In a meta-analysis of eighty-six studies investigating the effectiveness of second language pronunciation instruction, Lee, Jang, and Plonsky (2015) concluded that research into the use of technology in pronunciation teaching showed an overall smaller effect than human-based teaching, thus raising questions about its comparative value. In addition, approaches to CAPT are often not well-grounded in terms of their linkage to the knowledge base and theory characterizing pronunciation—though with notable exceptions like Cauldwell's (2012) *Cool Speech* and Thomson's (2012) *English Accent Coach*—as Thomson (2011) points out for CAPT in relation to L2 accent, its perceptual source in L1 transfer, and the differing goals of intelligibility and achievement of nativelike pronunciation. When research on a certain teaching approach or technology is not grounded in theory and a valid conception of the phenomenon under investigation, coupled with strong research methodology, it is difficult to interpret observed effects and to give convincing explanations for positive or negative findings. Thus, the basis for applying the teaching approach or technology in instruction is suspect.

Overall, Lee et al.'s (2015) aggregated findings suggest a strong positive effect of pronunciation teaching. A similar conclusion was drawn by Thomson and Derwing (2015), who conducted a narrative review of many of the same studies. Lee et al.'s (2015) analysis showed the benefits of pronunciation instruction particularly for instruction (a) over longer periods of time, (b) which provide feedback, and (c) which have more controlled outcome measures (p. 345). Their analysis also supported previous findings (e.g., Li, 2010; Plonsky, 2011; Spada, 1997) that laboratory-based pronunciation instruction may produce stronger effects than classroom-based teaching, although both are effective. Regarding pedagogical content, their analysis demonstrates that pronunciation teaching can be effective across a wide range of phonological features and that a greater effect is achieved by teaching both segmental and suprasegmental features (Lee et al., 2015, p. 361).

However, Lee et al. (2015) also highlight some important concerns with research into pronunciation teaching effectiveness. In particular, they point out that the small sample size of many studies and the lack of diversity (in terms of age, L1 background, and target language) reduces the validity (and specifically, the generalizability) of such research. Similarly, the general lack of delayed post-tests makes it difficult to determine the long-term effects of instruction. They also claim that many studies provide very little information about the actual approaches used to teach pronunciation: “Sorely missing from our results is a more fine-grained analysis of the effects of different types of pedagogic practice” (Lee et al., 2015, p. 361). Finally, they refer to the lack of focus in the group of studies reviewed on some aspects of pronunciation, including articulation, rhythm, linking, and stress. Despite such criticisms, the overall conclusion of Lee et al.’s (2015) research is that pronunciation teaching has a positive impact and both L2 speech perception and production can be improved by instruction. Although more research on the effects of instruction is needed, particularly more longitudinal and experimental studies, such research confirms that teaching pronunciation can be beneficial.

Another area of pronunciation teaching that has received considerable attention is form-focused instruction (FFI). A number of studies have shown the effectiveness in general of FFI, whether implicit or explicit in Ellis’ (2009) sense (see Chap. 4) and whether focus-on-form feedback in the course of negotiation for meaning (Long, 2015, p. 451) or other kinds of focus-on-forms (Chap. 4), including a wide-ranging study by Spada (1997) and three meta-analytical studies carried out since then (Goo, Granena, Yilmaz, & Novella, 2015; Norris & Ortega, 2000; Spada & Tomita, 2010). In the most recent decade, there have been a number of studies on the effectiveness of various kinds of FFI with a specific focus on pronunciation. An FFI effectiveness study specifically focused on pronunciation is that of Sato and Lyster (2012), who found in a study of Japanese student learners of English that corrective feedback (in the form of either prompts or recasts) was more effective for improving accuracy and fluency than peer interaction alone. A study by Saito and Lyster (2012) found that corrective feedback (in the form of recasts) to learners’ mispronunciations of English /ɹ/ during communicative tasks, after a 4-hour session of explicit perceptual input and practice on English /ɹ/, improved

Japanese learners' production of this feature, while the form-focused input without such corrective feedback did not. Saito's (2011, 2013) research on Japanese learners of English examined corrective feedback with and without provision of explicit phonetic information in the form of repeated modelling of exaggerated pronunciation of English together with presentation of a rule on the articulation of this phoneme contrasted with meaning-oriented input without any FFI. Saito found that the corrective feedback together with the explicit phonetic input had larger and more generalized effects than corrective feedback alone.

Some effectiveness studies have attempted to show the value of pronunciation instruction focused on segmental or suprasegmental features. A study by Derwing, Munro, and Wiebe (1998) demonstrated that instruction on suprasegmental aspects of pronunciation yielded better comprehensibility scores than instruction focused only on segmental aspects. A more recent experimental study by Saito and Saito (2016) resulted in significant improvements in comprehensibility and specific aspects of prosody which were the focus of a program that included pedagogical input and modeling, practice, and corrective feedback during meaningful interaction and that was geared specifically for Japanese university student EFL learners at a beginning level of proficiency. In another investigation, Gordon et al. (2013) compared the effectiveness of explicit segmental and suprasegmental FFI on comprehensibility scores of intermediate-level university ESL learners. While all groups were exposed to the same communicative methodology and sequence, one group was exposed to explicit phonetic instruction and analysis on segmental features (pairs of commonly confused vowel phonemes); one group was exposed to explicit phonetic instruction and analysis on suprasegmental features (stress, rhythm, linking, reductions); and one group received no explicit instruction on these features. Significant improvement in comprehensibility was seen only for the group with explicit focus on suprasegmentals, perhaps in part because, as the researchers note, the instruction on those features involved communication in discourse contexts whereas the instruction in segmentals was focused on simple lexical contrasts. Although it might be observed that a discourse-level context is a natural one for prosody, this difference in focus raises issues of the equivalence of the treatments, as Gordon et al. (2013) imply in their discussion of findings.

Longitudinal studies of pronunciation teaching, as in the case of other areas of teaching research, are few in number. However, two studies which have examined the long-term effects of pronunciation teaching are those of Couper (2006) and Sardegna (2011). Couper's (2006) study focused on students' insertion (epenthesis) or non-insertion of schwa in consonant clusters. Compared to a control group, he measured changes in pronunciation immediately after explicit instruction and then in a post-test 12 weeks later. As he reports, "Dramatic gains were achieved: the average error rate dropped from 19.9% to 5.5% in the immediate post-test, and rose slightly to 7.5% in the delayed post-test" (Couper, 2006, p. 46, Abstract). Sardegna (2011) investigated the accuracy of phonological linking between words (*is awesome*) and within words (e.g., *adept*) by a group of 38 ITAs when reading aloud and found significant improvements after four months' instruction. Additional tests 5 months later and then again after another 9 months showed that students who were available for later testing had maintained improvements in their accuracy.

A summary of key findings in the effectiveness of pronunciation teaching is provided in Table 8.2.

Table 8.2 Summary of key findings of pronunciation effectiveness research

Finding	Focus	Study
Effectiveness of pronunciation teaching	Meta-analysis Comparison with laboratory-based teaching	Lee et al. (2015); Li (2010); Plonsky (2011); Thomson and Derwing (2015)
Perceptual training improves production	HVPT (high variability perceptual/phonetic training)	Bradlow et al. (1997); Logan et al. (1991); Thomson (2011); Wong (2012)
Effectiveness of teaching suprasegmental aspects of pronunciation	Comprehensibility	Derwing et al. (1998); Gordon et al. (2013); Saito and Saito (2016)
Long-term effectiveness of pronunciation teaching	Epenthesis Linking	Couper (2006) Sardegna (2011)
Effectiveness of corrective feedback and form-focused instruction (FFI)	Overall accuracy and fluency Accuracy of English /ɪ/	Sato and Lyster (2012) Saito (2011, 2013); Saito and Lyster (2012)

In summary, although there has been a considerable amount of research into the effectiveness of pronunciation teaching showing the overall benefits of pronunciation instruction, there is a lack of rigorous, intervention-based studies employing control groups and delayed post-tests. The dearth of such studies is not surprising, given the practical difficulties of conducting carefully controlled research with real learners and in real classrooms. Also, according to Derwing and Munro (2015), many of the studies that have been done in this area tend to relate effectiveness to a change in accentedness, as judged by native speaker standard accent, rather than increased intelligibility; and, as they point out, “a change in accent does not ensure an improvement in communication effectiveness” (p. 97). However, we do note that a number of instructional intervention studies focused on suprasegmental features (e.g., Derwing et al., 1998; Gordon et al., 2013; Saito & Saito, 2016) use comprehensibility as the key measure of effectiveness. The appropriateness of comprehensibility as a measure of effectiveness in intervention studies focused on suprasegmental features is supported by research examining the linguistic correlates of native speaker judgements of the comprehensibility of the extemporaneous speech of adult L2 learners (Saito, Trofimovich, & Isaacs, 2016, 2017). Finally, as with other areas of pronunciation research, it is not evident how much of these findings filter through to pronunciation teaching practice.

English as an International Language or Lingua Franca

The significance of English as an international language (EIL) has long been recognized (e.g., Crystal, 1997; Kachru, 1985; Smith, 1976). Yet even as there is a growing awareness of the importance of pronunciation in maintaining successful communication in international contexts, there has also been much debate regarding the nature and role of English as an international or world or language. In the last twenty years, research into the use of English as a lingua franca (ELF) has reinforced not only the critical part played by pronunciation between L2 English speakers with different L1s, but also the ways in which the pronunciation priorities in ELF contexts may be evolving (Jenkins, 2000; Pickering, 2006).

It has also been argued that the rapid growth in the use of English as a global language, with the result that L2 speakers now outnumber L1 speakers (Graddol, 2006), has serious implications for English language teaching (ELT) policy, pedagogy, and assessment. As we saw in Chap. 3, key among these is a reconsideration of pronunciation models and targets, and the proposal that the pronunciation goal in lingua franca contexts should be “international intelligibility” with allowance for different L2 accents and that priorities should be limited to teaching a restricted pronunciation syllabus, which Jenkins (2000) refers to as the “lingua franca core” (LFC). A somewhat different orientation is a view of EIL as a relatively “neutral” or “universal” form of English devoid of specific features of L2 accent.

The recommendation of a discrete set of phonological features to use in prioritizing pronunciation teaching is an example of application of research to pedagogic practice. However, despite the appeal of the approach, the LFC has led to considerable debate among teachers and researchers, as we reviewed in Chap. 3. From both a teaching and a research point of view, several issues have been raised regarding the LFC. For instance, the omission of word stress from the core has been questioned (Walker, 2010; Walker & Zoghbor, 2015) and the limited sample size of the study on which the recommended core is based suggests a need for further, larger scale studies in different contexts, such as Deterding’s (2013) research into L2 English speakers in Southeast Asian contexts. Another issue is the lack of nuance in the LFC, which can be interpreted as suggesting that all L2 speakers can communicate successfully with other L2 speakers with this limited repertoire of English pronunciation. This may well be the case in some routinized, everyday encounters, such as ordering food at a café or booking into a hotel, but does not acknowledge the frequency of serious misunderstandings and communication breakdowns in both L2–L2 and L2–L1 encounters in many types of real-world settings of English language use, notably in the medical profession (e.g., as documented by Labov & Hanau, 2011) and contexts of business outsourcing in call centers (e.g., as documented by Lockwood et al., 2008; Tomalin, 2010; Yau, 2010). As highlighted in Chap. 7, awareness of such misunderstandings can make teachers reluctant to adopt the reduced Lingua Franca Core for teaching.

It is in fact not evident how aware or accepting pronunciation teachers are of the concept of the LFC, or whether they are in fact using it to design or select pronunciation components for their syllabus. Jenkins' (2005) own research into L2 learners' and teachers' perceptions and views on ELF reveals a largely negative attitude towards non-native English accents and a strong attachment to native speaker models and norms in pronunciation teaching. This conservatism has been reflected in some other studies (e.g., Mazlum, 2015; Scheuer, 2005; Schwartz, 2005). Recently, researchers have attempted to engage and work collaboratively with ELF-aware teachers to re-examine current methodology and practices (Cogo & Dewey, 2012; Zoghbor, 2011). However, many researchers refer to teachers' skepticism and unease with regard to the pedagogical implications of ELF. As Mazlum (2015, p. 100) observes, although they are generally accepted in outer or expanding circle contexts (Kachru, 1985), LFC-informed, EIL-based pronunciation approaches seem to be resisted by teachers in expanding circles.

An important area of debate with the LFC is its pedagogical scope: it can be argued that the narrow scope of the LFC risks underestimating the importance of situation-specific segmental and prosodic variants in constructing and interpreting meaning at many levels, the strong connection of pronunciation to identity and presentation of self, the reality that phonologically simplified speech is subject to stereotyping, and the difficulty of elaborating simplified phonology once it becomes entrenched, as we pointed out in Chap. 3. Advocates of teaching based on the LFC can rightly argue its strength and value in focusing on intelligibility and away from native speaker models for pronunciation. At the same time, critics can argue that this focus means important features of communicative competence in pronunciation are not addressed, thus leading some to react against teaching based on the LFC and to argue a need for expert-level (though not necessarily native speaker) models and broader goals than intelligibility for pronunciation competence.

The issues raised by ELF research present obvious potential challenges for English language teaching. As discussed in Chap. 3, many teachers are hesitant about adopting ELF recommendations, given some of the issues raised and the relatively small research base underlying this approach to pronunciation, in addition to the hesitations and objections of students. As Jenkins (2012) states:

ELF research findings question many long held beliefs about what and how English should be taught and tested: hitherto there has been little discussion of what this means for ELT professionals. And this has led to a feeling of unease and insecurity among them, as tends to happen whenever existing language standards or pedagogies are challenged. (p. 492)

Jenkins (2012) also claims that despite the changing use and roles of English, little has been done to relate these changes to ELT and that teaching, testing, and materials are still predominantly based on native speaker norms of correctness and appropriateness (p. 493). The gap between ELT research and theory and ELT practice is reflected in a “hands-off” attitude on the part of ELF researchers in relation to ELT practitioners. As Jenkins (2012) states, ELF researchers “do not believe it is our place to tell teachers what to do, but that it is for ELT practitioners to decide whether/to what extent ELF is relevant to their learners in their context” (p. 492). Research results are, after all, only suggestive, and the teacher plays a critical role in deciding the relevancy of specific research findings and teaching approaches to specific learners and contexts.

Intelligibility

Many of the current applied phonology textbooks (e.g., Collins & Mees, 2013; Cruttenden, 2014; Roach, 2009) pay explicit attention to the notion of intelligibility, addressing how this concept is being reconsidered to reflect the changing uses of English (Jenkins, 2000). Terms such as “comfortable intelligibility” (Abercrombie, 1949; Kenworthy, 1987), referring to speech which can be understood with little or no conscious effort on the part of the listener, and “international intelligibility” (Smith & Nelson, 1985) have been used to refer to intelligibility among speakers in international contexts. The notion of intelligibility is also being reconsidered in the light of the fact that intelligibility requires a two-way effort between speaker and listener (Munro &

Derwing, 1995; Smith & Nelson, 1985), as emphasized in Chap. 1. Jenkins (2012) also raises the question of “intelligible to whom?”, claiming that whereas intelligibility has traditionally been viewed from the standpoint of the L1 listener, in EIL contexts it can also be considered from the point of view of both L1 and L2 listeners, and that in ELF contexts it is important to consider mutual intelligibility between L2 speakers and listeners.

One of the difficulties with the concept of intelligibility is that it can be interpreted differently and indeed there are several different definitions of the term. Smith and Nelson’s (1985) tripartite definition describes intelligibility as the ability of the listener to recognize the speaker’s individual words or utterances, comprehensibility as the listener’s ability to understand the meaning of the word or utterance in its given context, and interpretability as the ability of the listener to understand the speaker’s intentions behind the word or utterance. Munro, Derwing, and Morton (2006) define intelligibility as “the extent to which a speaker’s utterance is actually understood” (p. 112), while Derwing and Munro (2015, p. 5) refer to intelligibility as “the degree of match between a speaker’s intended message and the listener’s comprehension” and define comprehensibility as the amount of effort required by the listener to understand a message. Derwing and Munro’s definition is similar to one stated in information processing terms which Smith and Nelson make reference to, as the extent to which a listener is able to receive a message as it was intended to be sent and to decode its elements. While these definitions of intelligibility are all roughly equivalent, they differ in detail and imply different measures for assessing intelligibility.

Our own discussion of intelligibility and comprehensibility in Chap. 1 draws on these definitions in a way that seeks to avoid referencing them both to comprehension. We think there is value in maintaining a clearer separation between these constructs and also a basis for doing so. In our view, intelligibility is based on a “threshold level” of pronunciation skill or competence in terms of speaker production and listener perception that is required for any communication to take place. This threshold level

may vary from one to another speaker and audience/addressee since intelligibility is always relative to a specific pairing and interaction of speaker and listener. Comprehensibility is a less distinctly phonological competence or judgement, incorporating segmental and prosodic features of pronunciation as well as temporal and lexicogrammatical aspects of L2 speech (Saito, Trofimovich, & Isaacs, 2016, 2017).

A radically different view of intelligibility is that of Rajagopalan (2010), who takes a political stance, suggesting that “intelligible” is merely “an evaluatory adjective like beautiful [or] ugly” (p. 468), seemingly refuting or ignoring empirical evidence that speakers can sometimes produce messages which are not understandable by most or all listeners. Although the way he states it is questionable, Rajagopalan nonetheless makes the important point that intelligibility is in the eye of the beholder—or rather, more accurately, the ear (and brain) of the listener—requiring processing of input that is subject to judgement.

The lack of unanimity and consistency in defining and measuring intelligibility interferes with the ability of pronunciation teachers, learners, and assessors to make appropriate application of intelligibility to language performance. There has been similar lack of consensus regarding the assessment of intelligibility (see Chap. 6), not only in terms of how to sample it (reading aloud or free speech), but also what measures or criteria to use (impressionistic or objective, analytic or holistic) and how to assess it (human or machine rating). Regarding the latter issue, for instance, there is some debate about the relative value of human and computer rating, with some noting that computer-based measurement is more accurate and reliable than human rating and avoids subjectivity (Kang, Rubin, & Pickering, 2010), while others argue that since intelligibility is an aspect of interaction and involves subjective judgement, it can only be assessed by a human listener and not by an acoustic analysis of speech (Derwing & Munro, 2005, p. 382). However, we note that creation of a database of utterances of a set passage by a wide range of speakers that has been assessed by human judges for intelligibility and then analyzed to discover the judges’ rating criteria, and/or for the types and degrees of divergence from judged intelligibility within the dataset, can be the basis for developing quantitative proxy measures of intelligibility which might be implemented on computer.

Another topic which is prompting considerable discussion and research is the relationship between intelligibility and accent. As discussed in several chapters in this book (see especially Chaps. 1, 3, and 7 for relevant discussion), studies have shown that there is not a straightforward correlation between accentedness and intelligibility and that it is possible for a speaker to have what is perceived as a “strong accent” and be yet be clearly intelligible (Derwing & Munro, 1997; Munro & Derwing, 1995). However, listener factors, such as attitudes towards various speaker groups and accent familiarity, can have a significant impact on how they judge the characteristics and the intelligibility of a speaker, as discussed in Chaps. 1 and 6 (see also Harding, 2011). Differential familiarity by listeners with specific accents has been noted as producing “construct-irrelevant variance” and low inter- and intra-rater reliability in language assessment (Browne & Fulcher, 2017; see also Chap. 6).

We have seen in recent decades a growing body of language learning and teaching theory and research which has proposed that intelligibility rather than nativelike proficiency should be the primary goal for language learners, based on the empirical evidence that few adult learners achieve native-like pronunciation in a second language (Flege et al., 1995) and also that intelligibility and accentedness are complex concepts which are partially independent (Derwing & Munro, 1997; Munro & Derwing, 1995) and have different implications for communicative impact and effectiveness. Awareness of research in these areas might help convince language teachers and learners to aim for more achievable or specifically relevant and purposeful goals. The notion of intelligibility as a legitimate learner goal is now relatively well established, thanks partly to the dissemination of a significant body of applied linguistic research focused on this notion, though issues and debates remain in this area of research and practice. There are also still many significant issues surrounding intelligibility and accentedness in relation to learner preferences, needs, and communicative effectiveness in different circumstances, as we reviewed here and in Chaps. 3 and 7.

Intonation and Prosody

There is a long history of research into prosody, especially intonation and stress, in L1 and L2 speech in many languages including English. However, as we have pointed out before regarding phonological research

in general, much of this has not filtered through to language teaching, despite the existence of some excellent books connecting intonation theory and teaching practice (e.g., Brazil, Coulthard, & Johns, 1980; Chun, 2002; Wells, 2006; Wennerstrom, 2001). Research has confirmed the complexity and multi-dimensionality of intonation (Bolinger, 1986, 1989; Halliday & Greaves, 2008; Ladd, 2008; Pierrehumbert, 1980) and has also revealed much about its structure and functions. Definitions of intonation vary but generally refer to changes in pitch level and patterning, or speech melody, above the word level. Intonation and prosody more generally, including choice of pitch contour or tone (e.g., a rise or a fall), degrees of stress, and phrasing (i.e., coarticulation, linking, and the placement of prosodic boundaries), can be exploited to convey a wide range of meaning, including:

- relations to context (*textual meaning*);
- logical sequence (*ideational meaning*); and
- social relations (*interpersonal meaning*). (Halliday & Greaves, 2008)

The basic structural components of intonation and prosody are now fairly firmly established, although there are some differences in descriptive traditions. There is increasing awareness of the multiple functions of intonation, such as for information structuring and for conveying semantic, pragmatic, interactional, and attitudinal meaning. However, there is still considerable debate regarding how intonation works in language; which functions or elements are marked within a specific language and which ones are basic or universal; and where the dividing line is between intonation, paralinguistics, and voice quality (see discussion in Chap. 1).

Today, most pronunciation teacher reference books include substantial sections on intonation and related features of prosody (e.g., Brown, 2014; Celce-Murcia, Goodwin, Brinton, & Griner, 2010; Collin & Mees, 2013; Grant, 2014; Reed & Levis, 2015; Rogerson-Revell, 2011). Many pronunciation textbooks also contain exercises and activities on intonation and other prosodic or suprasegmental features, some of which are engaging and informative, such as *Clear Speech* (Gilbert, 2012), *Speaking Clearly* (Rogerson & Gilbert, 1990), and the *Headway Pronunciation* (1999) series. However, many teaching materials concentrate on a fairly limited

range of intonational or prosodic functions, typically emphasizing the connection between intonation and (a) grammatical form (e.g., tone choice or sentence contour for *yes/no* and tag questions); (b) attitudes and emotion (e.g., rising tone used to convey surprise or sarcasm); and (c) information focus or structuring (e.g., *How did you stop smoking?*). The third category, information focus and structuring, relates to the process of signaling what is important in discourse by emphasizing new (unknown, foregrounded, or contrastive) information and de-emphasizing given (known, backgrounded, or non-contrastive) information. The importance of highlighting information, known by many terms—including “prominence,” “accentuation,” “sentence focus,” “sentence stress,” “nuclear stress,” “contrastive stress,” or “emphasis”—is well documented. The pedagogical focus on the grammatical and attitudinal functions of intonation reflects early work on English intonation, such as that of O’Connor and Arnold (1961), which suggested that individual tones or contours (e.g., fall, rise, fall-rise) convey specific meanings in conjunction with certain sentence types or attitudes. However, this view does not reflect research findings stemming, for example, from Discourse Analysis and Conversation Analysis that show that tone choices do not have a one-to-one correspondence with grammatical or attitudinal categories and that the meaning does not lie in a specific melodic pattern per se but in the context within which it is used (Cauldwell & Hewings, 1996; Crystal, 1969).

Approaches to teaching intonation which emphasize grammatical and emotional/attitudinal aspects tend to underplay the role of intonation in structuring discourse managing interactions, conveying pragmatic meaning, and cueing metamesages (Chap. 1). It is these informational, discursive, and pragmatic functions which are increasingly seen as underpinning the intonational system in language, whereby pitch variations in an utterance can signal to the listener how to unpack information and how to interpret a message—for instance, highlighting/de-emphasizing information, signaling topic change, ceding and taking conversational turns, and expressing interpersonal rapport. This is not to deny the attitudinal and grammatical aspects of intonation but to stress how the various intonational components interrelate under a broad discursive/pragmatic umbrella. As Levis and Wichmann (2015) state, “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'" (p. 149). Although listeners can make some initial inferences of pragmatic meaning and force of utterance based on the physical characteristics of a person's intonation and other prosodic features of voice—such as high or low; rising or falling; and unvarying or varying in pitch, stress, and/or volume—understanding the meaning of these features is a complex interpretive process involving not only prosodic characteristics and patterning, but also the entire communicational setting, including the specific choice of words and grammatical structures, facial expression and other bodily gestures, and prior communication with and knowledge of the speaker (if any).

Research suggests that L2 learners may not be aware of the pragmatic implications of intonation to signal speaker intent (Low, 2006; Pennington & Ellis, 2000; Reed & Michaud, 2015). For instance, in Low's (2006) pedagogically oriented research, cited as a case study in Chap. 3, it was determined that L2 English speakers in Singapore neither recognized nor used the conventional British English intonation pattern signaling given information. As another example, in a pilot study reported by Reed and Michaud (2015, pp. 459–461), Reed found that university-level learners of English in the United States did not attend to marked intonation and sentence focus when trying to interpret an utterance. When given the audio statement, *The teacher didn't grade your papers*, and asked to answer the question, *Have the papers been graded?*, learners initially responded negatively, ignoring the implication of the stress on *teacher*. Reed and Michaud (2015) suggest that it is therefore not enough to get students to mimic intonation patterns and assume they have learned intonation; they need to become metacognitively aware of how to interpret intonational meaning. They claim that it is this level of metacognitive awareness which is the missing link between intonation research and teaching practice and that "explicit instruction into the pragmatic norms surrounding intonation (telling learners that intonation can trump the words in an utterance and signal specific alternate meanings) is essential" (Reed & Michaud, 2015, p. 464).

Part of the problem may be that some teachers have limited awareness of the multiple functions of intonation and of current research in the field. As with pronunciation more generally, this situation may be due in part to inadequate teacher education for pronunciation compounded by lack of opportunities, encouragement, or motivation for teachers to

update their skills and knowledge in this area. This can result in a two-fold problem, 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” (pp. 72–73).

Another issue is that knowledge of which patterns and functions of intonation are universal or common across languages and which are less common (“marked”) or specific to particular languages and communicative functions is as yet quite limited. This has consequences for pedagogy: however interesting, or even easy to teach, an aspect of intonation is (e.g., intonational phrasing, turn holding), if the form and function of intonation is the same across languages, there seems little point in teaching it. Some functions that are phonologically “marked” in English, such as contrastive stress (e.g., *When did he leave?—He didn't leave.*), may well be worthwhile to teach: contrastive stress will be new to some learners and seems quite easily learnable (Pennington & Ellis, 2000).

It seems that the process of accentuating prominent syllables involves the same parameters across languages, although with different weightings of specific prosodic features, such as duration, amplitude, and pitch (Gordan & Applebaum, 2010). However, the process of deaccenting unimportant information is not universal, and many languages, as well as some varieties of English (e.g., Indian English, Caribbean English, and some East Asian varieties), do not follow the pattern of General British English (GB) or General American English (GA). Teaching variations in how information is signaled as given or backgrounded may therefore be important pedagogically, for awareness-raising and for communication with L1 English speakers (as examined in Chap. 3 in the Low, 2006, case study). This raises a further issue regarding the importance of teaching intonation in different learning contexts. It seems that some aspects of intonation are likely to be more important in ESL or EFL contexts, involving communication with L1 English speakers and listeners—for instance, the use of tone to convey attitude or pragmatic intent, or the use of contrastive stress, as illustrated above. In ELF contexts, however, means other than intonation are often used to indicate emphasis, focus, and speaker intent (Hirst & Di Cristo, 1998), such as a change in word order or word choice,

or the use of final attitudinal particles rather than contrastive stress. For example the meaning of English “Is that your bag?” might be rendered in French as *Est ce a vous ce sac?*, with marked placement of the possessive *a vous* indicating the emphasis.

As Roach (2009) writes, “I have witnessed many occasions when foreigners have unintentionally caused misunderstanding or even offence in speaking to a native English speaker, but can remember only a few occasions when this could be attributed to ‘using the wrong intonation’; most such cases have involved native speakers of different varieties of English, rather than learners of English” (p. 151). We note that “using the wrong intonation” has been documented in the example given in Chap. 1 as causing significant communication problems for Indian and Pakistani speakers of English working at London’s Heathrow Airport (Gumperz, 1982), and for Indian and Filipino speakers of English as documented in the examples given in Chap. 7 of problems in call center encounters (Friginal, 2007; Lockwood et al., 2008; Tomalin, 2010; Yau, 2010). However, these examples support Roach’s informal observation as long as these speakers are considered not “foreigners” but proficient and fluent speakers of “outer-circle” varieties of English. In contexts where “international intelligibility” is the aim, both Jenkins (2000) and Cruttenden (2014) claim that there is sufficient similarity across languages (e.g., in intonational phrasing and the use of rising and falling tones) to ensure intelligibility; and Cruttenden concludes that in such contexts, “no effort is needed to acquire the intonation patterns of L1 English” (p. 345). However, further empirical research is required to support or refute such a claim.

It is clear that intonation is a complex, multifunctional aspect of phonology which can be described in terms of some general features (e.g., in the treatment of intonation by Brazil, 1997) while also being describable in rich detail and complexity (as in its treatment by Bolinger, 1986, 1989). The treatment of intonation by Halliday and Greaves (2008) perhaps stands at a middle position in terms of generality and complexity in classifying differences in tone patterns into the three categories of ideational, textual, and interpersonal meaning. There is still much to be researched and learned about how intonation varies in context and relates to other areas of language and communi-

cation. The gap between theoretical and applied linguistic research in this field is starting to shrink, with a growing body of literature aimed at raising teachers' awareness of intonation form and meaning (Wichman, 2000; Levis & Wichmann, 2015) and suggesting pedagogical approaches to teaching intonation (Cauldwell & Hewings, 1996; Gilbert, 2008; Rogerson-Revell, 2012). There is nonetheless still research that needs to be done regarding which aspects of intonation and other aspects of prosody are important for intelligibility, comprehensibility, interpretability, fluency, interpersonal dynamics, and impact, and which teaching and learning methods are most effective in acquiring the various components of intonation and prosody more generally.

Some Common Myths and Misconceptions in Pronunciation Teaching and Learning

In the previous section we pointed out some of the key areas where a gap seems to exist between pronunciation research and practice. In relation to pronunciation teaching, this disconnect can lead to a largely intuition-based rather than evidence-based approach to pronunciation pedagogy (Derwing & Munro, 2015). Such a disconnect can give rise to, or perpetuate, beliefs and misconceptions which have little or no basis in empirical reality (Rogerson-Revell, 2011, 2015). For instance, beliefs such as “accent is the primary cause of unintelligibility”; “all *wh*-questions have a falling tone”; or “English is a stress-timed language” are common. The perpetuation of such misconceptions is unsurprising, given that many teachers feel underprepared to teach pronunciation (Foote et al., 2011) and may rely on teaching in ways they themselves were taught (Murphy, 2011). One way that research findings can help is by enabling teachers to revisit long-established beliefs or approaches to pronunciation teaching. In some cases this may mean revising or rejecting such beliefs in light of new findings, while in others it may result in reviving notions that were previously considered “old-fashioned” (see Rogerson-Revell, 2015). Some examples are considered below.

The Nature of Pronunciation

A common misconception by language teachers, students, and indeed most people is that pronunciation is largely a matter of standard language phonemics (or phonics), possibly coupled with some limited prosodic features such as strong/weak stress in words and larger units and falling/rising intonation in declarative and imperative versus interrogative sentences. As we have described and illustrated throughout this book, there is much more to pronunciation than this, and so much more that deserves attention in language instruction and testing in school and workplace contexts. At the segmental level, the necessary knowledge base includes not only phonemes but phonetic or contextual variants, regional as well as social, and the ways in which coarticulatory processes and speech style affect the pronunciation of segmental phonemes. At the suprasegmental level, the necessary knowledge base includes stress and accentuation in relation to lexical, grammatical, and informational structure as well as a broader knowledge of prosodic features and how these contribute to interactional dynamics, the development of information, and communicational effectiveness. A much richer and comprehensive understanding of pronunciation and its interconnections with intelligibility, presentation of self, and communicative impact is now available, as can be seen in the work of specialists in applied phonology over the last two decades and as reviewed in this book, with the inevitable effect of this knowledge gradually becoming disseminated to the various sites of pronunciation practice and research.

There is however a long way to go before the accumulating knowledge base finds its way into language teaching, testing, and other sites of pronunciation practice and research. All those involved with pronunciation teaching, testing, training, and/or research need to make efforts to expand their horizons and update their knowledge of pronunciation in its contextual and social aspects and with attention to the different kinds of contributions it makes to communication. This includes the most macro level, for the speaker, of creating first impressions and framing communication and, for the listener, of jump-starting the process of interpreting all that a person is communicating, down to the most micro level of articulating or decoding individual sounds as intelligible words. Language materials developers need to bring what they are disseminating

to teachers up to date and more in alignment with current understandings and knowledge about pronunciation, while language testers need to revisit pronunciation as a construct and reconsider language assessment in light of pronunciation dynamics and complexity.

Tone and Sentence Types

An example of a pronunciation myth driving teaching practice is the meaning ascribed to utterance-final tones, in traditional pedagogical approaches that present “rules” which associate sentence types with intonation patterns. It is commonly claimed that *wh*-questions and statements terminate in a falling tone and *yes/no* questions end in a rising tone. Research using conversational and corpus-based data of both American English (Levis, 1999) and British English (Cauldwell & Hewings, 1996; Grabe, Kochanski, & Coleman, 2005) has shown that these “rules” are commonly not followed, as has been understood for some time (Bolinger, 1986, 1989), suggesting that such an approach to teaching tones does not necessarily reflect the realities of natural speech. As Cruttenden (2014) states, “despite what is often stated in textbooks on English language teaching, both rises (usually low rises) and falls (usually high falls) occur frequently on *yes/no*-interrogatives and *wh*-interrogatives” (p. 325). Data from the Grabe et al. (2005) investigation, for example, show a great deal of prosodic variation within and between “varieties of English spoken in Cambridge, Newcastle and Belfast, [for which] questions and statements can be accompanied by a wide range of intonation patterns” (p. 331). As they conclude:

[A]uthors of textbooks cannot find it easy to decide which intonation pattern they should suggest for a particular utterance type. They may wish to point out that the pattern they provide is the frequent or typical option.... [L]earners of English need to be prepared for extensive variation in the intonation they might hear from native speakers, within and across dialects. Most of all, they need to be aware that variation in the southern ‘standard’¹ is as high or higher than in northern varieties of English spoken in the British Isles. In other words, the standard variety is no more uniform than non-standard varieties. (ibid.)

More discourse-oriented descriptions suggest that a rising tone indicates openness or uncertainty and a falling tone, finality or certainty (Brazil, 1997; Cruttenden, 1997/1986; Wichmann, 2000) and that these sorts of general or iconic meanings associated with rising and falling pitch may be the basis for the grammatical associations with interrogative and declarative sentence types, with *wh*-questions seeming to fall between the two types. These and further studies involving the detailed analysis of natural conversational speech continue to reveal the complexities of tone choice in interaction, including signaling speaker status and dominance or power (Bolinger, 1986, 1989; Brazil, 1997; Gussenhoven, 2004) and pragmatic intent (Levis & Wichmann, 2015). Such research highlights the importance of context in interpreting intonational meaning and the need for caution in making categorical matches between tones or intonation and sentence types, however convenient for teaching.

Rhythm

Most pronunciation teachers and teaching material rightly highlight the importance of rhythm in learning pronunciation and point out that the distinctive rhythm of English is related to its feature of strong and weak stress. Typically, an explanation is given based on the theory that the world's languages are divided into two (or three) types, those that have a "syllable-timed" rhythm (e.g., French, Spanish, Telugu, and Yoruba) and those that have a "stress-timed" rhythm (e.g., English, German, Russian, and Arabic). Mandarin, as opposed to Cantonese, is sometimes categorized as stress-timed or put in an "in-between" category, and Japanese is sometimes considered to belong to a third category of either a "pitch-timed" or a "mora-timed" language.² The distinction between stress-timed and syllable-timed languages is based on the notion that "[t]here is considerable variation in syllable length in a language spoken with stress-timed rhythm whereas in a language spoken with a syllable-timed rhythm the syllables tend to be equal in length" (Abercrombie, 1967, p. 98). As often claimed in pronunciation textbooks, the idea is that for stress-timed languages, the stresses occur at regular intervals, and "the unstressed syllables are squeezed into the time available, with the result that they may become very short" (Roach, 2001, p. 36).

However intuitively appealing this notion is, and despite its perpetuation in pronunciation textbooks and teacher education courses, there is not a strong base of research to support it. Indeed, many researchers have refuted the concept of stress-timing (Bolinger, 1986; Cauldwell, 2002; Dauer, 1983; Roach, 1982, 2001). Roach (1982) carried out a range of measurements to test the two features identified by Abercrombie (1967) as differentiating stress-timed and syllable-timed languages, with the following conclusions:

Firstly, ...there is no language which is totally syllable-timed or totally stress-timed—all languages display both sorts of timing; languages will, however, differ in which type of timing predominates. Secondly, different types of timing will be exhibited by the same speaker on different occasions and in different contexts.... Finally, the stress-timed/syllable-timed distinction seems at the present to depend mainly on the intuitions of speakers of various Germanic languages all of which are said to be stress-timed.... (p. 78)

Roach (2001) notes that the squeezing or shortening of weakly stressed syllables between equally spaced stressed syllables “is only found in a style of speech (slow, emphatic) where the rhythm is strong, and in ordinary conversational speech it is much harder to make a convincing case for this **isochronous** rhythm (where the time intervals between stressed syllables are equal)...” (p. 37, emphasis in original).

This is not to deny the importance of rhythm in communication or rhythmic differences between languages, individual speakers, and types of speech, as Roach (2001, p. 37) emphasizes. For example, formal, planned language, such as a sermon or speech, has a more regular rhythm than spontaneous talk, and English and other Germanic languages have less regular rhythm than Spanish, French, or Cantonese, which have less complex syllable types and make less use of distinctions between stress levels from one syllable to the next. However, accentuation or stress at word, phrase, and sentence level is what characterizes the rhythm of English: in particular, the distinctively strong contrast between stressed and unstressed syllables and the use of reduced vowels in unstressed syllables (Deterding, 2001). In addition to lexical and grammatical structure, speech tempo is also driven by information focus,

since non-focal information—that which the speaker does not stress informationally—will tend to be destressed phonologically as well (e.g., by being produced comparatively rapidly).

Rhythm-based practice (e.g., using poetry or other metrical material) may aid memorization and development of motor skills and add fun to learning, but pedagogically the danger is that learners conclude that English should be spoken with the sort of consistent rhythm which they might have been taught through limericks and rhyming poems. Nevertheless, the rhythmic patterning of real speech is a worthwhile topic for instruction, and the notion of highly varying and often alternating strong and weak syllables can be useful for conveying the importance of rhythm in relation to accentuation or stress for intelligibility in English (Rogerson-Revell, 2011), as well as in relation to communicative context, purpose, and impact. As Roach (2001) points out:

[R]hythm is useful to us in communicating: it helps us to find our way through the confusing stream of continuous speech, enabling us to divide speech into word or other units, to signal changes between topic or speaker, and to spot which items in the message are most important. (p. 37)

Thus, rhythm, particularly as it occurs and varies across different levels of language (syllable, word, phrase, clause, and multi-clause units) and types of communication, is a fruitful topic for pronunciation teaching and a much richer one than an emphasis on stress-timing would seem to imply.

Minimal Pairs

Minimal pairs, in which the lexical meaning of a word can be changed by replacing a single phoneme (e.g., *thin/tin*), are ubiquitous in pronunciation teaching materials and form the basis of much pronunciation teaching at segmental level. Minimal-pair activities generally focus learners' attention on sounds in isolation, often pronounced in an exaggerated or hyperarticulated way. Much of minimal-pair work involves decontextualized production or discrimination, or communicatively artificial contextualization since in very many cases the minimal-pair words being targeted would rarely occur in the same contexts. It is not clear that this

kind of focus has much transfer to language use in authentic communication, and teaching based on minimal pairs draws learning time and attention away from contextual variants and the coarticulatory linking, levelling, and loss processes involved in producing speech in real time (Hieke, 1985). In addition, minimal-pair activities are often designed without any consideration of the importance of the contrast exemplified, largely based on the assumption of many teachers and materials writers that all minimal pair contrasts are equally significant and therefore valuable for teaching. However, this assumption is not borne out by research, which suggests that some contrasts are considerably more important than others and “only a small number of those minimal pairs currently in pronunciation books should be used in pronunciation teaching” (Levis & Cortes, 2008, p. 204).

Misunderstandings in natural speech are rarely a result of minimal pair confusions with no other factors, and Levis and Cortes (2008, p. 205) claim that there is little evidence for the assumption that the mispronunciation of one sound will be enough to irretrievably harm understanding or create serious interference or breakdown in communication. Even when a word intended by a speaker as one member of a minimal pair is pronounced in such a way that it is heard by a listener as the other member of that minimal pair—that is, as the “wrong word”—communication often continues, with the error receiving little attention beyond a brief notice that the speaker probably mispronounced or misspoke—a common occurrence in both spontaneous and prepared speech. Outside some special situations such as legal contexts, it is probably rare for communicators to interpret what is being said word by word, and in the full depth and detail that might be possible. Rather, much of communication goes on at a relatively general level of connotation and gist. In fact, even if a listener fails to deduce that a speaker produced the wrong member of a minimal pair and instead hears the word which was produced as the other member of the minimal pair, the listener may not recognize any misunderstanding but rather search for an interpretation for the word spoken which will make sense. Communicators make great efforts to deduce meaning from context, in order to make sense of whatever they heard, or think they heard.

Pennington (2002, p. 440) illustrates a range of possibilities in the example of a misalignment between a speaker's production and a listener's perception of the word *block*, heard as *black*, ranging from perception of an error in pronunciation (i.e., mispronouncing the word *block* in a way that sounded like *black*) to a context-related interpretation in which *black* would make sense. Although such cases of minimal pair confusions are thought to be rare, they may be more common than we think but, for various reasons (e.g., expediency, tolerance of error, focus away from details of form), do not end up being flagged within the flow of conversation. Nevertheless, there can occasionally be situations where misperception of one word for another, as in the air traffic control and medical examples given in Chaps. 1 and 7 and the other medical examples mentioned in Case Study 2 (Labov & Hanau, 2011) in Chap. 3 are not a minor matter and can be literally a matter of life and death.

Minimal pair exercises can be useful to assess discrimination of sound contrasts in the target language and for micro-level listening practice, if contextualized, used with relatively common words of the same word class and with semantically plausible contrasts (e.g., *pork/fork*, *copy/coffee*). Brown (1988, p. 601) suggests that these requirements do not apply to the majority of minimal pairs in English and that the concept of functional load (discussed in Chap. 3) is useful in deciding which phonemes do the most work in discriminating words (e.g., /ɪ/ vs. /i:/, as in *live/leave*), and therefore are the most salient for teaching in terms of their functionality and cue value. Again, a closer link between research findings in this area and materials developers and teachers would be of great value.

Diphthongs

Again, most teaching materials and teachers consider the teaching of diphthongs as an essential part of teaching English vowels. However, this is an area of language where there is considerable difference between varieties and accents and also where pronunciation is evolving quite rapidly, thus an area where established practices and assumptions may need to be examined. Traditionally, phonological descriptions show that GB has eight diphthongs: /ɪə/ *here*; /eə/ *hair*; /ʊə/ *sure*;

/eɪ/ *bay*; /aɪ/ *buy*; /ɔɪ/ *boy*; /əʊ/ *boat*; and /aʊ/ *shout*. GA or North American English (NAE) is typically described as having five diphthongs: /aɪ/ *buy*, /aʊ/ *bough*, /ɔɪ/ *boy*, *bait* /eɪ/, and *boat* /oʊ/.³

Only the three closing (or “upgliding”) diphthongs—/aʊ/, /aɪ/, and /ɔɪ/—are common across all native speaker varieties. Many diphthongs either vary considerably across regional dialects or are changing across generations. In GB, /ɔɪ/, /ʊə/, and /eə/ are unstable; /ʊə/ is increasingly replaced by the long monophthong /ɔ:/ (e.g., *poor*, *sure*) and /eə/ by the long monophthong /ɛ:/ (e.g., *square*, *hair*). Cruttenden (2014) suggests that L2 learners of English can pronounce the diphthong /eɪ/ (*bait*) and /əʊ/ (*boat*) as pure vowels (that is, as [e:] and [o:]), which is common pronunciation in some parts of the United Kingdom, including Scotland and Northern England, as well as in many of the New Englishes which have developed outside the inner circle countries (see Pennington, 1996, pp. 105–111).

In a description of pronunciation features necessary for use in “International English” contexts, Cruttenden (2014) predicts that only two diphthongs will remain, /aɪ/ and /aʊ/. Similarly, Jenkins (2000) does not include any diphthongs or vowel quality contrasts in the LFC. Both researchers claim that vowel contrasts in general seem less crucial for intelligibility than consonant contrasts, particularly in ELF contexts, and the vowel length distinction is more important than vowel quality differences, because there is substantial variation among L1 English speakers. However, we note that North American English varieties are still prone to have diphthongs in all of the contexts mentioned, so shifting to monophthongs for most of those vowels will generally mark a speaker as speaking a non-standard or non-NAE variety.

This example illustrates another area where practitioners might benefit by knowing the results of research but where there is a lack of dissemination between researchers and practitioners. It also highlights the general lack of consideration given in teaching and learning to regional and social variation and to the dynamic, evolutionary nature of pronunciation. An obvious area is the choice of a pronunciation model or standard for L2 learners. We have considered previously how the increasing use of English as a lingua franca is changing the needs and priorities of some L2 speakers,

and a large body of research is emerging focusing on how language, including pronunciation itself, is evolving due to the predominant use of English among L2 speakers. At the same time, the pronunciation of L1 English speakers is evolving. For instance, in the United Kingdom, this is happening not only at the segmental level (as with diphthongs), but also in relation to features of connected speech, such as use of glottal stop instead of linking between a final stop and an initial vowel in the next word, as in “quite easy” [kwaɪ? 'i:zi] (Wells, 1982, p. 324), and with intonation, such as the increasing use of final high rising tone (HRT) in “Upspeak” (see Chaps. 1 and 2), typically by younger users (Bradford, 1997; Pennington, Lau, & Sachdev, 2011).

There seems to be little evidence as yet of such changes filtering through to most pronunciation teaching, textbooks, or teacher reference books (though see Pennington, 1996, which gives considerable attention to variation in different varieties of English), but some learners are aware that the pronunciation they have been taught originally seems old-fashioned (e.g., producing a vowel in *bad* which sounds more like *bed* to GB speakers) or does not correspond to what they hear from L1 English speakers. Teachers may not be able to rely on intuition to deal with such concerns and would benefit from knowledge updates from researchers and teacher educators.

Some Gaps in the Current Knowledge Base

(i) *Insufficient dissemination of research findings to teachers and other practitioners*

Despite a growing body of research in the field and recent calls by pronunciation specialists to reconsider goals, targets, and approaches to teaching pronunciation in light of the changing role and status of English as an international language and lingua franca in an increasingly plurilingual world (Jenkins, 2000; Seidlhofer, 2011; Sewell, 2013), little of this research seems to have informed the practices of language teachers, assessors, curriculum designers, or materials developers. The research disconnect is especially notable in relation to the meta-analytical effectiveness research

establishing the value of FFI (Goo et al., 2015; Norris & Ortega, 2000; Spada & Tomita, 2010) and to a lesser extent to the meta-analytic effectiveness research demonstrating the value of teaching not only segmental but also suprasegmental features (Lee et al., 2015). This disconnect is illustrated by the lack (until 2015⁴) of a dedicated journal applying phonological research to pronunciation teaching. Although there is a considerable body of L2 speech research published in well-established journals, Derwing and Munro (2005) suggest that such publications are not aimed at a language teaching audience and therefore these findings do not feed down into pronunciation teaching or teaching textbooks. Derwing and Munro (2005) claim that “this situation thus creates a twofold problem: relatively little published research on pronunciation teaching and very little reliance on the research that exists” (p. 383).

(ii) *Insufficient knowledge of effective pronunciation teaching methods and tools*

Research into pronunciation teaching per se is a relatively recent phenomenon with relatively few studies in the twentieth century as contrasted with the still-young twenty-first. In recent decades, more studies have been undertaken exploring various aspects of pronunciation instruction, although there is still a dearth of empirically grounded research to help teachers prioritize issues for the pronunciation classroom or to clarify the effectiveness of various approaches to pronunciation learning and teaching. There are still relatively few, well-designed intervention-based studies, particularly longitudinal studies with delayed post-tests, which are needed to provide rigorous evidence which can either support or refute current teaching priorities and practices.

(iii) *Insufficient clarity of key concepts*

Another issue that hampers communication between researchers and practitioners is the lack of clarity and consensus regarding some key concepts and terms in the field. Although much pronunciation-related research investigates concepts such as *intelligibility*, *fluency*, *proficiency*, *accent*, and *pronunciation* itself (see discussion in Chap. 6), there is

considerable variation in how such constructs and terms are defined and used. Commonly understood and agreed definitions are not only important for the development and replication of research but are also essential for the rigorous development and implementation of pronunciation assessments, particularly in high-stakes tests such as IELTS and TOEFL.

(iv) *Insufficient interactional, data-driven research*

Some aspects or contexts of pronunciation are difficult to research. For instance, we described earlier some of the complexities and constraints on conducting classroom-based research (e.g., ethical issues, gaining access, and recruiting participants). Research into intelligibility is similarly fraught with difficulties, such as how to collect and analyze valid speech samples, whether to use perception-based human raters (and whether L1 or L2 speakers) or instrumental acoustic measures. Partly because of such difficulties, a substantial amount of research involves the collection and analysis of short samples of decontextualized, monologic speech (either individual words or short phrases) and little contextualized, interaction-based data. The lack of naturalistic data is not surprising, given the problems of investigating intelligibility in authentic interactions (Coetzee-Van Roey, 2009; Rajadurai, 2007), where the authenticity of the data results in variation which makes it challenging to draw measurable and valid conclusions. It is also very difficult in naturalistic, interactive situations to assess understanding and intelligibility or the role of pronunciation in achieving it (Rogerson-Revell, 2014). Nevertheless, this is not to say that such data is not valuable. Interactive speech data can be very revealing in terms of enabling the observation of how participants react to phonological events in such contexts, whether in terms of mispronunciations which result in communication breakdown or which have no noticeable effect.

For instance, in Rogerson-Revell's (2008) research examining pronunciation in the context of international business meetings in English, both types of instance were observed. In one example, segmental errors resulted in a request for clarification between an L1 Spanish speaker and an L1 Dutch speaker and in another, an L1 French speaker with what was considered a "strong" L1 accent conversed with considerable fluency with other speakers and without any apparent misunderstandings (Rogerson-Revell, 2008).

Such data can be valuable for supporting or refuting claims such as that a strong accent impedes intelligibility. There are other examples of spoken language corpora being used for phonological analysis at segmental and suprasegmental levels (e.g., Brazil, 1997; Brazil et al., 1980; Cauldwell, 2002), but there could be a much greater focus on discourse-based research in this field.

(v) *Insufficient cross-fertilization among disciplines*

As we and others (e.g., Derwing & Munro, 2015; Levis, 2016) have pointed out, there is a substantial body of research relevant to L2 pronunciation in other disciplines: for instance, work in the speech sciences into articulation and remediation of pronunciation disorders, use of experimental methods such as eye tracking in psychology and fMRI in neurolinguistics to study the working of the brain during speech processing, and work in sociolinguistics examining pronunciation in relation to identity and social context. Although a considerable amount of this research is relevant to L2 phonology, much of it goes largely unread by pronunciation researchers and practitioners. Similarly, much of the research by applied linguists in areas such as intelligibility and language competence is unknown to speech scientists. This lack of cross-fertilization among disciplines not only leads to a danger of “reinventing the wheel,” but also restricts opportunities to broaden perspectives and conduct collaborative research.

(vi) *Insufficient basic and validation research on pronunciation assessment*

Pronunciation assessment has not for the most part developed as a focal area of applied practice, other than for the assessment of pronunciation aptitude as a key aspect of language aptitude (see Chaps. 2 and 6), but has rather been included as a component of the overall assessment of speaking. As pronunciation receives more attention in its own right, this leads naturally to critical examination of assessment practices. Active areas of research involving pronunciation assessment that researchers agree require further attention include the rating scales and criteria used for scoring pronunciation, the complex interaction of raters' characteristics upon how they score different learner groups and aspects of pronuncia-

tion, and ongoing reliability and validity research on machine testing and scoring of pronunciation. In addition to these areas of active research, there are a number of issues requiring attention from researchers in relation to test development and validity. A large area needing basic research and validation studies is the construct of pronunciation and its components, including both perceptual and social competencies, and how these can be incorporated in assessment of pronunciation within overall speaking assessment and as an autonomous area of assessment.

(vii) *Insufficient research on pronunciation in non-educational contexts*

The majority of research on pronunciation is focused on education and instructional contexts. Yet, as we have shown in this book, specifically in Chap. 7 but also in many of the examples in Chap. 1, pronunciation is a topic of importance in many other contexts which have, however, received considerably little research attention. These include many kinds of business and professional contexts (e.g., medicine, call centers) in which pronunciation affects communication and assessed performance. Yet the amount of research on pronunciation which has been conducted in these non-educational contexts is so far quite small. The need is great and still growing, with increasing flows of people around the globe and increasing interactions of people from different language backgrounds that continue to increase the need for L2 pronunciation skills.

(viii) *Insufficient research on pronunciation in languages other than English*

Given the dominance of English as an international language and global lingua franca, it is perhaps not unexpected that the amount of research on English language pronunciation teaching and other aspects of applied practice and research (e.g., forensics, specific business and professional contexts) far outweighs that in all other languages. While much that can be learned from the studies on English will have direct applicability to other languages, there are likely to be significant differences as well, based on different needs and circumstances related to the learning of specific languages. It would therefore be of great value to expand research agendas for pronunciation in other languages. Such research can

directly benefit the teaching of English by offering up-to-date knowledge for purposes of comparison and contrast, and for establishing the pronunciation characteristics and knowledge that forms the baseline or starting point of different learner groups. Besides their value for these comparative purposes and also for testing the generalizability of the methods and results of studies carried out on English, an important benefit of such research is to focus more attention on languages other than English as targets for teaching and learning. In addition to expanding research on other languages as the target of instruction or other kinds of applied practice, there is a need for more research on English language pronunciation learning, teaching, and other aspects of applied practice for different L2 learner groups.

Key Areas for Future Research

A logical place to start in our consideration of future research is with the teaching approaches for pronunciation that have been recommended by scholars and pronunciation specialists based on previous research and/or their knowledge of the teaching and learning of pronunciation. Of value would be classroom-based studies on teaching approaches such as those recommended by Low (2006) that compare the prosodic patterns of different varieties and/or compare prosodic and non-prosodic ways of conveying meaning, including how teachers' and learners' beliefs, attitudes, awareness, and approaches to pronunciation change as a result of such instruction. Also of value would be systematic study of the approaches advocated by Couper (2015) based on Cognitive Phonology that focus on critical listening to build awareness of cues to phoneme category boundaries, building a shared metalanguage for pronunciation with students, communicative practice, and effective feedback. In addition, comparison of different classes in terms of the pronunciation metalanguage that is evolved in each case could provide valuable input for teacher education and student-oriented teaching materials for pronunciation.

Studies might be conducted of teaching sequences that move systematically from a focus on form to communicative practice such as that advocated in Pennington (1996, pp. 224–243) or in Celce-Murcia

et al. (2010), following the lead of Gordon et al.'s (2013) research, or from a top-down focus on prosody and vocal setting to a bottom-up focus on segmental phonology such as advocated in Pennington (1989). Such systematic and sequenced curricula, which are grounded in theory and have been advocated for some time, have hardly been researched. A number of different types of research on sequenced pronunciation curricula would be valuable, including:

- comparative curriculum studies, such as one course that focuses on pronunciation only as it arises in communication versus another that treats pronunciation in one of the recommended sequences or some variant of these;
- detailed case study or ethnography of a pronunciation-focused class, examining both teacher and learner attitudes, behaviors, and outcomes;
- longitudinal study of learners' changing metalinguistic awareness of pronunciation and their changing perceptual and productive abilities after each step in the pronunciation sequence.

Gordon et al.'s (2013) discussion of the contrasting contexts of prosodic and segmental orientations to pronunciation instruction suggests that comparisons between the two in terms of instructional effects or effectiveness are not easy to draw. This is because a pedagogical focus on suprasegmental or prosodic aspects of pronunciation, unless restricted to word stress or unless taught in a strictly phonetic way, will bring in matters of social meaning that would not necessarily be addressed by a segmental focus, which would typically address lexical meaning such as in minimal pairs. The differences in context and focus can be seen to make the two different treatments essentially incomparable for the purposes of drawing conclusions from research studies.⁵ In an attempt to address equivalence of these two different treatments, important areas for future research comparing segmental versus prosodic orientations for instruction include comparing them in ways that attempt to offer more similar kinds of instruction across these two very different features of pronunciation. These might involve, for example contrasting treatments based on strictly phonetic kinds of input and perhaps also devising research to

compare segmental and suprasegmental approaches at a lexical level, for example, teaching multisyllabic words based on the specific articulation of the vowels and consonants, or based on the overall stress and intonation pattern without any reference to segmentals. It might also be possible to devise studies of more socially oriented approaches to teaching segmental phonology, based, for example, on real data or constructed contexts in which listeners express their lack of understanding or give feedback on impact or other impressions to L2 speakers based on their segmental errors or confusions, to explore whether pronunciation instruction tied to social meaning will be more effective than that tied strictly to denotative meaning.

Studies which continue to explore connections between accentedness and intelligibility and comprehensibility for different audiences and purposes can help to inform instruction, as can further studies of the pronunciation attitudes and goals of different learner populations in regard to accent, intelligibility, and communicative purpose and effectiveness. In this vein, more studies of learner attitudes, pronunciation learning, and communicative competence and effectiveness within an EIL or ELF-oriented curriculum, such as one focused on the Lingua Franca Core advocated by Jenkins (2000), are needed as a basis of pronunciation practice for different learner populations.

A focus for future research studies that is in step with current language learning and teaching theory is the provision and comparison of different types of input-rich environments. As discussed in Chap. 4, these might involve extensive listening and flooding learners with input in the beginning stages of language learning as a way for the learner to focus on and gain familiarity with how the L2 sounds, and to start building a perceptual model for the L2 *ab initio*, from the beginning and from scratch, as a form of implicit learning based strictly on L2 input in natural communicative contexts, rather than from L1 transfer and explicit learning. In particular, we would like to see studies that systematically delay the introduction of any kind of written support or explicit teaching of language structure for a period of weeks with the goal of imprinting the sound pattern of the language as the foundation for all other aspects of language learning, in attempting to reduce the effects to L1 transfer and equivalence classification in the beginning stages of language learning.

There are in addition a number of active research frontiers that can be explored by different kinds of methodologies. These include pedagogically focused research evaluating technologies that are currently being used or might profitably be used in pronunciation instruction. This is an area of practice that is in rapid development and in which applied research and theory have some catching up to do, as compared to teaching techniques that have emerged from synergies between (i) educational and language teaching theory and (ii) SLA research and theory. Other research frontiers that are attracting a lot of attention in relation to pronunciation are the cognitive basis of pronunciation in relation to memory and language aptitude and the connections of pronunciation to personality, identity, and social orientation. Given their apparently substantial contribution to pronunciation performance, as reviewed in Chap. 2, these seem to be important phenomena to explore in order to develop theoretical and practical linkages to teaching. Researching these areas and their contributions to individual pronunciation performance and differences in pronunciation outcomes may spur innovations in teaching to emulate the strategies and behaviors of the “good pronunciation learner” and to deliver individualized instruction of new types and in new ways. Finally, we suggest that it might be useful to design studies that are aimed at investigating commonalities rather than differences between L1 and L2 acquisition, and between language learning in childhood and adulthood, including in contexts of childhood bilingualism, as a way to gain new perspectives on the processes involved that might give direction for instruction.

Concluding Remarks

There appears to have been a strong resurgence of interest in L2 pronunciation research in recent years, which is both reassuring and promising. The number of pronunciation-related articles is increasing and covering a wider range of research journals (Levis, 2016). In addition, starting in 2015, the first subject-specific journal of L2 pronunciation, *The Journal of Second Language Pronunciation*, began publication. Nevertheless, as we have pointed out in this chapter, there are still areas in need of a stronger research focus, in pedagogy as well as in a number of other areas of research and practice involving pronunciation.

As we suggested earlier, more carefully controlled intervention-based studies are needed. If rigorously carried out, such studies will provide empirically grounded evidence which can translate into recommendations to teachers and other practitioners. However, laboratory-based studies do not necessarily provide the only or best evidence, and a single small-scale study can provide misleading results in terms of generalizable, large-scale or long-term effects. There is a need for more classroom-based investigations, including with teachers and researchers themselves as participants. Teachers are well situated to be involved in classroom-based research, especially for conducting observational studies or action research in their own classrooms. More extensive or in-depth attention to pronunciation in teacher education may encourage such research, and research collaborations between classroom practitioners and university researchers can be especially valid and productive. As well as more classroom-based research and intervention studies exploring pronunciation teaching in different contexts, there is a need to broaden the research scope to include more discourse and corpus-based studies of authentic speech in order to investigate pronunciation in a variety of contexts and types of interactions.

An understanding of key issues and relevant research can help practitioners diagnose errors, prioritize pronunciation content, and select relevant teaching, training, and assessment methodologies. For example, knowing about HVPT or other perceptual or form-focused instructional techniques may encourage a teacher or a workplace trainer to try this approach to learning segmental phonemes rather than relying on teaching minimal pairs, a technique far removed from teaching pronunciation in communicative contexts. As another example, knowing that the “rule” that *wh*-questions and statements terminate in a falling tone and *yes/no* questions end in a rising tone does not always apply in context may lead a teacher or other practitioner to explore a broader, more contextualized approach to working on tones and intonation. Awareness of research which suggests that suprasegmental aspects of pronunciation appear to cause more comprehension difficulties for L1 listeners while segmental errors seem to be more critical for L2 listeners (Derwing & Munro, 2005; Jenkins, 2000), or that prosody is at least as important as segmental aspects of pronunciation for conveying communicative affect and metamesages in

many kinds of interactions could influence what priority a teacher gives to segmental or suprasegmental aspects of pronunciation, depending on the specific learners' needs.

Although there is a growing amount of interesting and relevant research in the field, there still appears to be a lack of synergy between research and pedagogy, and between research and practice more generally. Teachers have limited awareness of significant findings which can influence their decisions regarding pronunciation teaching content, goals and priorities, and approaches and methods. Other practitioners, including language testers and trainers in workplace environments, also frequently lack knowledge of research findings that are relevant to their work. On the other side of the two-way trading relation between research and practice, researchers may have a limited view of pronunciation focused on their specific area of investigation and not necessarily informed by the realities of pronunciation in classroom or other real-world settings, thus limiting the applicability, generalizability, and validity of their findings. The synergy between pronunciation research and practice needs to be a two-way process with knowledge and dialogue flowing in both directions. Relevant research findings should have an impact on pedagogy and changes in pedagogy should lead to new research questions and directions.

Applied phonologists and teacher educators have an important role to play in describing research findings in a form that teachers can understand and appreciate, and in making recommendations for application of research findings in classrooms and other settings of practice that teachers can make use of directly and that textbook writers can elaborate in curriculum and course designs and materials. This points up a need for research "mediators" and "translators" who interact with, and as, both teachers and researchers, thereby helping to ensure an ongoing and robust trading relationship between research and pedagogy. As these mediators disseminate information on both sides, they also help to connect the two informational and professional bodies in a bidirectional interaction in which the findings of research have an impact on pedagogy, and the changing and variable contexts of pedagogy provide continuing practical and theoretical validation for research findings, even as they also drive further investigations.

As in the case of validation of tests, which we explored in Chap. 6, validation of specific practices and research findings is an ongoing process that involves considerations of how context impacts outcomes. The development of a body of reliable and valid research findings and practices for pronunciation will therefore only come through continuing dialogue and interaction between researchers and practitioners. In addition, progress in the field of pronunciation or applied phonology, as in other areas of applied knowledge, can benefit from continuing cross-fertilization of both theoretical and applied bodies of knowledge, ranging from language learning and educational theory and methods, to sociolinguistics and cognitive psychology. Considering the very large and diverse population of L1 and L2 learners and speakers of English who are the audience of pronunciation research and practice, we suggest that the best outcomes will result from synthesizing knowledge from all of these different areas that connect to pronunciation.

Notes

1. Grabe et al. (2005) equate this to the Cambridge accent.
2. A mora is a segment the length of a short vowel; long vowels are two moras in length. Vowel length is phonemic in Japanese.
3. The latter two vowels are sometimes classified as tense vowels, especially in respect to non-standard dialects where they maybe monophthongal; but for standard varieties of NAE they are both upgliding diphthongs (see discussion in Pennington, 1996, pp. 105–118).
4. The *Journal of Second Language Pronunciation* was established in 2015 (<https://benjamins.com/#catalog/journals/jslp.1.1/main>).
5. It could be reflected that the difference between the two instructional approaches is not just one between apples and oranges—which would be an appropriate type of difference for making comparisons—but more like the difference between apple pie and orange juice—that is, too different to make any kind of meaningful comparisons.

References

- Abercrombie, D. (1949). Teaching pronunciation. *ELT Journal*, 3(5), 113–122. <https://doi.org/10.1093/elt/III.5.113>
- Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh: Edinburgh University Press.
- Bolinger, D. (1986). *Intonation and its parts: Melody in spoken English*. Stanford, CA: Stanford University Press.
- Bolinger, D. (1989). *Intonation and its uses: Melody in grammar and discourse*. Stanford, CA: Stanford University Press.
- Bradford, B. (1997). Upspeak in British English. *English Today*, 13(3), 29–36. <https://doi.org/10.1017/S0266078400009810>
- Bradlow, A. R., Pisoni, D., Akahane-Yamada, R., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America*, 101(4), 2299–2310. <https://doi.org/10.3758/BF03206911>
- Brazil, D. (1997). *The communicative value of intonation in English* (2nd ed.). Cambridge: Cambridge University Press.
- Brazil, D., Coulthard, M., & Johns, C. (1980). *Discourse intonation and language teaching*. London: Longman.
- Brown, A. (1988). Functional load and the teaching of pronunciation. *TESOL Quarterly*, 22(4), 593–606. <https://doi.org/10.2307/3587258>
- Brown, A. (1991). *Teaching pronunciation: A book of readings*. London and New York: Routledge.
- Brown, A. (2014). *Pronunciation and phonetics: A practical guide for English language teachers*. New York: Routledge.
- Browne, K., & Fulcher, G. (2017). Pronunciation and intelligibility in assessing spoken fluency. In T. Isaacs & P. Trofimovich (Eds.), *Second language pronunciation assessment: Interdisciplinary perspectives* (pp. 37–53). Bristol, UK and Blue Ridge Summit, PA: Multilingual Matters.
- Catford, J. C. (1987). Phonetics and the teaching of pronunciation: A systemic description of English phonology. In J. Morley (Ed.), *Current perspectives on pronunciation: Practices anchored in theory* (pp. 87–100). Alexandria, VA: TESOL.
- Cauldwell, R. (2002). The functional irrhythmicity of spontaneous speech: A discourse view of speech rhythms. *Apples*, 2(1), 1–24. Retrieved July 30, 2017, from http://apples.jyu.fi/article_files/cauldwell_2002.pdf
- Cauldwell, R. (2012). *Cool speech*, app. Retrieved July 30, 2017, from <http://www.speechinaction.org/cool-speech-2/>

- Cauldwell, R. & Hewings, M. (1996). Intonation rules in ELT textbooks. *English Language Teaching Journal*, 50(4), 327–334. <https://doi.org/10.1093/elt/50.4.327>
- Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M., with Griner, B. (2010). *Teaching pronunciation: A course book and reference guide* (2nd ed.). New York: Cambridge University Press.
- Chun, D. M. (2002). *Discourse intonation in L2: From theory and research to practice*. Amsterdam: John Benjamins.
- Coetzee-Van Roey, S. (2009). Intelligibility and perceptions of English proficiency. *World Englishes*, 28(1), 15–34. <https://doi.org/10.1111/j.1467-971X.2008.01567.x>
- Cogo, A., & Dewey, M. (2012). *Analysing English as a Lingua Franca: A corpus-driven investigation*. London: Continuum/Bloomsbury.
- Collins, B., & Mees, I. (2013). *Practical phonetics and phonology: A resource book for students* (3rd ed.). London and New York: Routledge.
- Couper, G. (2006). The short and long-term effects of pronunciation instruction. *Prospect*, 21, 46–66 Retrieved July 22, 2017, from http://www.ameprc.mq.edu.au/docs/prospect_journal/volume_21_no_1/21_1_3_Couper.pdf
- 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. <https://doi.org/10.1080/09658416.2011.570347>
- Couper, G. (2015). Applying theories of language and learning to teaching pronunciation. In M. Reed & J. M. Levis (Eds.), *Handbook of English pronunciation* (pp. 413–432). New York: John Wiley & Sons.
- Cowie, C. (2007). The accents of outsourcing: The meanings of “neutral” in the Indian call center industry. *World Englishes*, 26(3), 316–330. <https://doi.org/10.1111/j.1467-971X.2007.00511.x>
- Cruttenden, A. (1997/1986). *Intonation*. Cambridge, UK: Cambridge University Press.
- Cruttenden, A. (2014). *Gimson's pronunciation of English* (8th ed.). London: Routledge.
- Crystal, D. (1969). *Prosodic systems and intonation in English*. Cambridge: University Press.
- Crystal, D. (1997). *English as a global language*. New York: Cambridge University Press.
- Cutler, C. (2014). Accentedness, “passing” and crossing. In J. M. Levis & A. Moyer (Eds.), *Social dynamics in second language accent* (pp. 145–167). Boston: DeGruyter Mouton.

- Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, 11(1), 51–62. Retrieved July 22, 2017, from <http://psycnet.apa.org/record/1983-29886-001>
- de Bot, K. (1983). Visual feedback of intonation I: Effectiveness and induced practice behavior. *Language and Speech*, 26(4), 331–350. <https://doi.org/10.1177/002383098302600402>
- De Bot, K., & Mailfert, K. (1982). The teaching of intonation: Fundamental research and classroom applications. *TESOL Quarterly*, 16(1), 71–77. <https://doi.org/10.2307/3586564>
- Derwing, T. M. (2003). What do ESL students say about their accents? *Canadian Modern Language Review*, 59(4), 545–564. <https://doi.org/10.3138/cmlr.59.4.547>
- Derwing, T. M., Diepenbroek, L., & Foote, J. (2012). How well do general-skills ESL textbooks address pronunciation? *TESL Canada Journal*, 30(1), 22–44. <https://doi.org/10.18806/tesl.v30i1.1124>
- Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition*, 19(1), 1–16. Retrieved January 10, 2018, from <https://www.cambridge.org/core/journals/studies-in-second-language-acquisition/issue/77E9B64BB6D1D3DC15A6523E69A42665>
- Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly*, 39(3), 379–397. <https://doi.org/10.2307/3588486>
- Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. Amsterdam: John Benjamins.
- Derwing, T. M., Munro, M. J., & Thomson, R. I. (2008). A longitudinal study of ESL learners' fluency and comprehensibility development. *Applied Linguistics*, 29(3), 359–380. <https://doi.org/10.1093/applin/amm041>
- Derwing, T. M., Munro, M. J., & Wiebe, G. E. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning*, 48(3), 393–410. <https://doi.org/10.1111/0023-8333.00047>
- Deterding, D. (2001). The measurement of rhythm: A comparison of Singapore and British English. *Journal of Phonetics*, 29(2), 217–230. <https://doi.org/10.1006/jpho.2001.0138>
- Deterding, D. (2013). *Misunderstandings in English as a Lingua Franca: An analysis of ELF interactions in South-East Asia*. Berlin: de Gruyter Mouton.
- Dörnyei, Z., & Csizér, K. (2002). Motivational dynamics in second language acquisition: Results of a longitudinal nationwide survey. *Applied Linguistics*, 23(4), 421–462. <https://doi.org/10.1093/applin/23.4.421>

- Ellis, R. (2009). Implicit and explicit learning, knowledge and instruction. In R. Ellis, S. Loewen, C. Elder, H. Reinders, R. Erlam, & R. Philp (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 3–25). Bristol, UK: Multilingual Matters/Channel View Publications.
- Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America*, 97(5), 3125–3134. <https://doi.org/10.1121/1.413041>
- Foote, J. A., Holtby, A. K., & Derwing, T. M. (2011). Survey of pronunciation teaching in adult ESL programs in Canada, 2010. *TESL Canada Journal*, 29(1), 1–22 181–196. <https://doi.org/10.18806/tesl.v29i1.1086>
- Foote, J. A., Trofimovich, P., Collins, L., & Urzúa, F. (2013). Pronunciation teaching practices in communicative second language classes. *Language Learning Journal*, 44(2), 181–196. <https://doi.org/10.1080/09571736.2013.784345>
- Friginal, E. (2007). Outsourced call centers and English in the Philippines. *World Englishes*, 26(3), 331–345. <https://doi.org/10.1111/j.1467-971X.2007.00512.x>
- Gilbert, J. B. (2008). *Teaching pronunciation using the prosody pyramid*. Cambridge: Cambridge University Press. Retrieved September 24, 2017, from <https://www.tesol.org/docs/default-source/new-resource-library/teaching-pronunciation-using-the-prosody-pyramid.pdf?sfvrsn=0>
- Gilbert, J. B. (2012). *Clear speech: Pronunciation and listening comprehension in North American English* (4th ed.). New York: Cambridge University Press.
- Gonet, W., Szpyra-Kozłowska, J., & Świąciński, R. (2010). Clashes with ashes. In E. Waniek-Klimczak (Ed.), *Issues in accents of English 2: Variability and norm* (pp. 213–230). Newcastle upon Tyne: Cambridge Scholars Publishing.
- Goo, J., Granena, G., Yilmaz, Y., & Novella, M. (2015). Implicit and explicit instruction in L2 learning. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 443–482). Amsterdam: John Benjamins.
- Gordon, J., Darcy, I., & Ewert, D. (2013). Pronunciation teaching and learning: Effects of explicit phonetic instruction in the L2 classroom. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 4th pronunciation in second language learning and teaching conference*, August 2012 (pp. 194–206). Ames, IA: Iowa State University.
- Gordon, M., & Applebaum, A. (2010). Acoustic correlates of stress in Turkish Kabardian. *Journal of the International Phonetic Association*, 40(1), 35–58. <https://doi.org/10.1017/S0025100309990259>
- Grabe, E., Kochanski, G., & Coleman, J. (2005). The intonation of native accent varieties of English in the British Isles: Potential for miscommunication? In K. Dziubalska-Kołodziejczyk & J. Przedlacka (Eds.), *English pronunciation models: A changing scene* (pp. 311–337). Bern: Peter Lang.

- Graddol, D. (2006). English next: Why global English may mean the end of “English as a Foreign Language”. The British Council. Retrieved September 11, 2017, from <https://englishagenda.britishcouncil.org/continuing-professional-development/cpd-researchers/english-next>
- Grant, L. (Ed.). (2014). *Pronunciation myths: Applying second language research to classroom teaching*. Ann Arbor: University of Michigan.
- Gumperz, J. J. (1982). *Discourse strategies*. New York: Cambridge University Press.
- Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge: Cambridge University Press.
- Halliday, M. A. K., & Greaves, W. S. (2008). *Intonation in the grammar of English*. London and Oakville, CA: Equinox.
- Harding, L. (2011). *Accent and listening assessment: A validation study of the use of speakers with L2 accents on an academic English listening test*. Frankfurt: Peter Lang.
- Henderson, A., Curnick, L., Frost, D., Kautzsch, A., Kirkova-Naskova, A., Levey, D., et al. (2015). The English pronunciation teaching in Europe survey: Factors inside and outside the classroom. In J. A. Mompean & J. Fouz-González (Eds.), *Investigating English pronunciation: Trends and directions* (pp. 260–292). Basingstoke, UK and New York, NY: Palgrave Macmillan.
- Henderson, A., Frost, D., Tergujeff, E., Kautzsch, A., Murphy, D., Kirkova-Naskova, A., et al. (2012). The English pronunciation teaching in Europe survey: Selected results. *Research in Language, 10*, 5–27. <https://doi.org/10.2478/v10015-011-0047-4>
- Henderson, A., & Jarosz, A. (2014). Desperately seeking a communicative approach: English pronunciation in a sample of French and Polish secondary school textbooks. *Research in Language, 12*(3), 261–277. <https://doi.org/10.2478/rela-2014-0015>
- Hieke, A. E. (1985). A componential approach to oral fluency evaluation. *Modern Language Journal, 69*(2), 135–142. <https://doi.org/10.1111/j.1540-4781.1985.tb01930.x>
- Hirst, D. J., & Di Cristo, A. (Eds.). (1998). *Intonation systems: A survey of twenty languages*. Cambridge: Cambridge University Press.
- Iverson, P., & Kuhl, P. K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of the Acoustical Society of America, 97*(1), 553–562. <https://doi.org/10.1121/1.412280>

- Jakobson, R. (1931). Prinzipien der historischen Phonologie. *Travaux du Cercle Linguistique de Prague*, 4, 247–267.
- Janicka, K., Kul, M., & Weckwerth, J. (2005). Polish students' attitudes to native English accents. In K. Dziubalska-Kořaczyk & J. Przedlacka (Eds.), *English pronunciation models: A changing scene* (pp. 251–292). Bern: Peter Lang.
- 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(3), 535–543. <https://doi.org/10.2307/3588493>
- Jenkins, J. (2012). English as a Lingua Franca from the classroom to the classroom. *ELT Journal*, 66(4), 486–494. <https://doi.org/10.1093/elt/ccs040>
- Kachru, B. B. (1985). Standards, codification and sociolinguistic realism: The English language in the outer circle. In R. Quirk & H. Widdowson (Eds.), *English in the world: Teaching and learning the language and literatures* (pp. 11–30). Cambridge: Cambridge University Press.
- Kang, O., Rubin, D., & Pickering, L. (2010). Suprasegmental measures of accentedness and judgments of language learner proficiency in oral English. *The Modern Language Journal*, 94(4), 554–566. <https://doi.org/10.1111/j.1540-4781.2010.01091.x>
- Kenworthy, J. (1987). *Teaching English pronunciation*. London: Longman.
- Labov, J., & Hanau, C. (2011). Pronunciation as life and death: Improving the communication skills of non-native English speaking pathologists. In B. J. Hoekje & S. M. Tipton (Eds.), *English language and the medical profession: Instructing and assessing the communication skills of international physicians* (pp. 261–285). Bingley, UK; Leiden and Boston: Emerald; Brill.
- Ladd, R. (2008). *Intonational phonology* (2nd ed.). Cambridge University Press.
- Leather, J. (1999). Second-language speech research: An introduction. *Language Learning*, 49(s1), 1–56. <https://doi.org/10.1111/0023-8333.49.s1.1>
- Lee, J., Jang, J., & Plonsky, L. (2015). The effectiveness of second language pronunciation instruction: A meta-analysis. *Applied Linguistics*, 36(3), 345–366. <https://doi.org/10.1093/applin/amu040>
- Levis, J. (2016). The interaction of research and pedagogy. *Journal of Second Language Pronunciation*, 2(1), 1–7. <https://doi.org/10.1075/jslp.2.1.001lev>
- Levis, J. M. (1999). Intonation in theory and practice, revisited. *TESOL Quarterly*, 33(1), 37–63. <https://doi.org/10.2307/3588190>

- Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, 39(3), 369–377. <https://doi.org/10.2307/3588485>
- Levis, J. M., & Cortes, V. (2008). Minimal pairs in spoken corpora: Implications for pronunciation assessment and teaching. In C. A. Chapelle, Y.-R. Chung, & J. Xu (Eds.), *Towards adaptive CALL: Natural language processing for diagnostic language assessment* (pp. 197–208). Ames, IA: Iowa State University.
- Levis, J. M., & Wichmann, A. (2015). English intonation—Form and meaning. In M. Reed & J. M. Levis (Eds.), *The handbook of English pronunciation* (pp. 139–155). Hoboken, NJ: John Wiley & Sons.
- Li, W. (2010). The nature of linguistic norms and their relevance to multilingual development. In M. Cruz-Ferreira (Ed.), *Multilingual norms* (pp. 397–404). Frankfurt am Main: Peter Lang.
- Lockwood, J., Forey, G., & Price, H. (2008). Englishes in the Philippine business processing outsourcing industry: Issues, opportunities and research. In M. L. S. Bautista & K. Bolton (Eds.), *Philippine English: Linguistic and literary perspectives* (pp. 157–172). Hong Kong: Hong Kong University Press.
- Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /a/ and /l/: A first report. *Journal of the Acoustical Society of America*, 89(2), 874–886. <https://doi.org/10.1121/1.1894649>
- Long, M. (2015). *Second language acquisition and task-based language teaching*. Malden, MA and Oxford, UK: Wiley Blackwell.
- 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. <https://doi.org/10.2307/40264306>
- MacIntyre, P. D., Clément, R., Dörnyei, Z., & Noels, K. A. (1998). Conceptualizing willingness to communicate in a L2: A situational model of L2 confidence and affiliation. *Modern Language Journal*, 82(4), 545–562. <https://doi.org/10.2307/330224>
- Mazlum, F. (2015). Pronunciation pedagogy in the expanding circle: Investigating Iranian teachers' attitudes to ELF and LFC. In *Proceedings of EPIP4, 4th international conference on English pronunciation: Issues & practices* (pp. 97–100). Prague: Institute of Phonetics, University of Prague.
- McCroskey, J. C., & Richmond, V. P. (1987). Willingness to communicate. In J. C. McCroskey & A. Daly (Eds.), *Personality and interpersonal communication* (pp. 129–156). Newbury Park, CA: Sage.
- Moyer, A. (2007). Do language attitudes determine accent? A study of bilinguals in the USA. *Journal of Multilingual and Multicultural Development*, 28(6), 502–518. <https://doi.org/10.2167/jmmd514.0>

- Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 45(1), 73–97. <https://doi.org/10.1111/j.1467-1770.1995.tb00963.x>
- Munro, M. J., & Derwing, T. M. (2006). The functional load principle in ESL pronunciation instruction: An exploratory study. *System*, 34(4), 520–531. <https://doi.org/10.1016/j.system.2006.09.004>
- Munro, M. J., Derwing, T. M., & Morton, S. L. (2006). The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition*, 28(1), 111–131. <https://doi.org/10.1017/S0272263106060049>
- Murphy, D. (2011). An investigation of English pronunciation teaching in Ireland: ELT in Ireland presents a number of interesting issues when it comes to the question of pronunciation. *English Today*, 27(4), 10–18. <https://doi.org/10.1017/S0266078411000484>
- Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, 50(3), 417–528. <https://doi.org/10.1111/0023-8333.00136>
- Nowacka, M. (2012). Questionnaire-based pronunciation studies: Italian, Spanish and Polish students' views on their English pronunciation. *Research in Language*, 10(1), 43–61. <https://doi.org/10.2478/v10015-011-0048-3>
- O'Connor, J. D., & Arnold, G. F. (1961). *Intonation of colloquial English*. London: Longman.
- Paunović, T., & Savić, M. (2008). Discourse intonation – Making it work. In S. Komar & U. Mozetič (Eds.), *ELOPE, V(1–2), As you write it: Issues in literature, language, and translation in the context of Europe in the 21st century* (pp. 57–75). Ljubljana: Slovene Association for the Study of English and Department of English, Faculty of Arts, University of Ljubljana. <https://doi.org/10.4312/elope.5.1-2.57-75>. Also available at <https://revije.ff.uni-lj.si/elope/article/view/3314/3007>
- Pennington, M. C. (1989). Teaching pronunciation from the top down. *RELC Journal*, 20(1), 20–38. <https://doi.org/10.1177/003368828902000103>
- Pennington, M. C. (1996). *Phonology in English language teaching: An international approach*. London and New York: Longman.
- Pennington, M. C. (2002). Real language phonology: Optimality Theory meets sociolinguistics. *Journal of Sociolinguistics*, 6(3), 418–448. <https://doi.org/10.1111/1467-9481.00195>
- Pennington, M. C., & Ellis, N. C. (2000). Cantonese speakers' memory for English sentences with prosodic cues. *Modern Language Journal*, 84(3), 372–389. <https://doi.org/10.1111/0026-7902.00075>

- Pennington, M. C., Lau, L., & Sachdev, I. (2011). Diversity in adoption of linguistic features of London English by Chinese and Bangladeshi adolescents. *Language Learning Journal*, 39(2), 177–199. <https://doi.org/10.1080/09571736.2011.573686>
- Pickering, L. (2006). Current research on intelligibility in English as a Lingua Franca. *Annual Review of Applied Linguistics*, 25, 219–233. <https://doi.org/10.1017/S0267190506000110>
- Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation*. PhD thesis, M.I.T.
- Plonsky, L. (2011). The effectiveness of second language strategy instruction: A meta-analysis. *Language Learning*, 61(4), 993–1038. <https://doi.org/10.1111/j.1467-9922.2011.00663.x>
- Rajadurai, J. (2007). Intelligibility studies: A consideration of empirical and ideological issues. *World Englishes*, 26(1), 87–98. <https://doi.org/10.1111/j.1467-971X.2007.00490.x>
- Rajagopalan, K. (2010). The soft ideological underbelly of the notion of intelligibility in discussions about ‘World Englishes’. *Applied Linguistics*, 31(3), 465–470. <https://doi.org/10.1093/applin/amq014>
- Reed, M., & Levis, J. M. (Eds.). (2015). *The handbook of English pronunciation*. West Sussex, UK: John Wiley & Sons, Inc.
- Reed, M., & Michaud, C. (2015). Intonation in research and practice: The importance of metacognition. In M. Reed & J. M. Levis (Eds.), *The handbook of English pronunciation* (pp. 454–470). Hoboken, NJ: John Wiley & Sons.
- Rindal, U. (2010). Constructing identity with L2: Pronunciation and attitudes among Norwegian learners of English. *Journal of Sociolinguistics*, 14(2), 240–261. <https://doi.org/10.1111/j.1467-9841.2010.00442.x>
- Rindal, U. (2013). *Meaning in English: L2 attitudes, choices and pronunciation in Norway*. Doctoral thesis, University of Oslo.
- Roach, P. (1982). On the distinction between ‘stress-timed’ and ‘syllable-timed’ languages. In D. Crystal (Ed.), *Linguistic controversies* (pp. 73–79). London: Arnold.
- Roach, P. (2001). *Phonetics*. Oxford introductions to language study. Oxford: Oxford University Press.
- Roach, P. (2009). *English phonetics and phonology* (4th ed.). Cambridge: Cambridge University Press.
- Rogerson, P., & Gilbert, J. B. (1990). *Speaking clearly*. Cambridge: Cambridge University Press.

- Rogerson-Revell, P. (2008). Participation and performance in international business meetings. *English for Specific Purposes*, 27, 338–360. <https://doi.org/10.1016/j.esp.2008.02.003>
- Rogerson-Revell, P. (2010). Can you spell that for us non-native speakers? Accommodation strategies in international business meetings. *Journal of Business Communication*, 47(4), 432–454. <http://journals.sagepub.com/doi/pdf/10.1177/0021943610377304>
- Rogerson-Revell, P. (2011). *English phonology and pronunciation teaching*. London: Bloomsbury.
- Rogerson-Revell, P. (2012). Can or should we teach intonation? *Speak Out! IATEFL Pronunciation SIG Newsletter*, 47, 16–20. Retrieved on July 22, 2017, from <https://pronsig.iatefl.org/Archive/SO%20Website%20PDFs/SO!%2047%20website.pdf>
- Rogerson-Revell, P. (2014). Pronunciation matters: Using English for international business communication. In R. van den Doel & L. Rupp (Eds.), *Pronunciation matters* (pp. 137–156). Amsterdam: VU Uitgeverij.
- Saito, K. (2011). *Effects of FFI on L2 phonological development of /ɹ/ by Japanese learners of English*. Doctoral dissertation, McGill University.
- Saito, K. (2013). Reexamining effects of form-focused instruction on L2 pronunciation development. *Studies in Second Language Acquisition*, 35(1), 1–29. <https://doi.org/10.1017/S0272263112000666>
- Saito, K. (2015). Experience effects on the development of late second language learners' oral proficiency. *Language Learning*, 65(3), 563–595. <https://doi.org/10.1111/lang.12120>
- Saito, K., & 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(2), 595–633. <https://doi.org/10.1017/S0272263112000356>
- Saito, Y., & Saito, K. (2016). Differential effects of instruction on the development of second language comprehensibility, word stress, rhythm, and intonation: The case of inexperienced Japanese EFL learners. *Language Teaching Research*, 21(5), 589–608. <https://doi.org/10.1177/1362168816643111>
- Saito, K., Trofimovich, P., & Isaacs, T. (2016). Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels. *Applied Psycholinguistics*, 37(2), 217–240. <https://doi.org/10.1017/S0142716414000502>
- Saito, K., Trofimovich, P., & Isaacs, T. (2017). Using listener judgements to investigate linguistic influences on L2 comprehensibility and accentedness: A

- validation and generalization study. *Applied Linguistics*, 38(4), 439–462. <https://doi.org/10.1093/applin/amv047>
- Sardegna, V. G. (2011). Pronunciation learning strategies that improve ESL learners' linking. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 2nd pronunciation in second language learning and teaching conference* (pp. 105–121). Ames, IA: Iowa State University.
- Sato, M., & Lyster, R. (2012). Peer interaction and corrective feedback for accuracy and fluency development: Monitoring, practicing, and proceduralization. *Studies in Second Language Acquisition*, 34(4), 591–626. <https://doi.org/10.1017/S0272263112000356>
- Scheuer, S. (2005). Why native speakers are (still) relevant. In K. Dziubalska-Kořaczyk & J. Przedlacka (Eds.), *English pronunciation models: A changing scene* (pp. 111–131). Bern: Peter Lang.
- Schwartz, G. (2005). The Lingua Franca core and the phonetics-phonology interface. In K. Dziubalska-Kořaczyk & J. Przedlacka (Eds.), *English pronunciation models: A changing scene* (pp. 152–177). Bern: Peter Lang.
- Seidlhofer, B. (2011). *Understanding English as a Lingua Franca*. Oxford: Oxford University Press.
- Sewell, A. (2013). English as a Lingua Franca: Ontology and ideology. *ELT Journal*, 67(1), 3–10. <https://doi.org/10.1093/elt/ccs061>
- Smith, L. E. (1976). English as an international auxiliary language. *RELC Journal*, 7(2), 38–42. <https://doi.org/10.1177/003368827600700205>
- Smith, L., & Nelson, C. (1985). International intelligibility of English: Directions and resources. *World Englishes*, 4(3), 333–342. <https://doi.org/10.1111/j.1467-971X.1985.tb00423.x>
- Spada, N. (1997). Form-focussed instruction and second language acquisition: A review of classroom and laboratory research. *Language Teaching*, 30(2), 73–87. <https://doi.org/10.1017/S0261444800012799>
- Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning*, 60(2), 263–308. <https://doi.org/10.1111/j.1467-9922.2010.00562.x>
- Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception & Psychophysics*, 36(2), 131–145. <https://doi.org/10.3758/BF03202673>
- Tergujeff, E. (2013). *English pronunciation teaching in Finland*. Jyväskylä, Finland: University of Jyväskylä. Retrieved September 23, 2017, from <https://tinyurl.com/nd624kz>. [Google Scholar](#)
- Tomalin, B. (2010). India rising: The need for two way training. In G. Forey & J. Lockwood (Eds.), *Globalization, communication and the workplace* (pp. 172–189). London: Continuum.

- Thomson, R. I. (2011). Computer assisted pronunciation training: Targeting second language vowels: Perception improves pronunciation. *CALICO Journal*, 28(3), 744–765. <https://doi.org/10.11139/cj.28.3.744-765>
- Thomson, R. I. (2012). *English accent coach*. Retrieved July 30, 2017, from <https://www.englishaccentcoach.com/>
- Thomson, R. I., & Derwing, T. M. (2015). The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics*, 36(3), 326–344. <https://doi.org/10.1093/applin/amu076>
- Underhill, A. (2012). Proprioception and physicality: *The physicality of pronunciation & proprioception*. Retrieved September 24, 2017, from <http://www.adrianunderhill.com/2012/08/28/proprioception-and-pronunciation/>
- 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. (2010). *Teaching the pronunciation of English as a Lingua Franca*. Oxford: Oxford University Press.
- Walker, R., & Zoghbor, W. (2015). The pronunciation of English as a Lingua Franca. In M. Reed & J. M. Levis (Eds.), *The handbook of English pronunciation*. Oxford: Wiley-Blackwell.
- Wells, J. C. (1982). *Accents of English 2: The British Isles*. Cambridge: Cambridge University Press.
- Wells, J. (2006). *English intonation: An introduction*. Cambridge: Cambridge University Press.
- Wennerstrom, A. K. (2001). *The music of everyday speech: Prosody and discourse analysis*. New York: Oxford University Press.
- Wichmann, A. (2000). *Intonation in text and discourse*. Harlow, UK: Longman.
- Wong, J. W. S. (2012). Training the perception and production of English /e/ and /æ/ of Cantonese ESL Learners: A comparison of low vs. high variability phonetic training. In *Proceedings of the 14th Australasian International Conference on Speech Science and Technology (SST 2012)*, Sydney, Australia (pp. 37–40).
- Yau, J. N. W. (2010). Call centre discourse: Graduation in relation to voice quality and attitudinal profile. In G. Forey & J. Lockwood (Eds.), *Globalization, communication and the workplace* (pp. 107–124). London: Continuum.
- Zoghbor, W. S. (2011). Teaching the pronunciation of English as a Lingua Franca: Reducing skepticism and increasing practicality. *International Journal of Humanities and Social Science*, 1(8), 285–288. Retrieved September 22, 2017, from http://www.ijhssnet.com/journals/Vol._1_No._8;_July_2011/35.pdf

Author Index¹

A

Abercrombie, D., 422, 434, 435
Abrahamsson, N., 78, 295
Abreu, M., 393
Ackermann, J., 110, 337
Acton, W., 89
Adams, M. J., 385
Adelstein, A., 61, 100n2
Adjémian, C., 77
Adolphs, S., 352, 356
Agger-Gupta, N., 352
Akahane-Yamada, R., 199, 414
Albertson, K., 268
Alfi-Shabta, I., 78
Allwright, R. L., 152
Alsabaan, M., 254
Al-Saraj, T. M., 87
Altenberg, E. P., 76, 77, 148
Altinsel, Z., 177

Altman, J., 245, 372
Anderson-Hsieh, J., 188, 304
Applebaum, A., 429
Aslin, R., 65
Aston, G., 210
Atkins, S., 357
Atkinson, J. M., 361, 362
Aziz-Zadeh, L., 98

B

Baddeley, A. D., 60, 383
Bailey, K., 10
Bailey, K. M., 177, 178
Baker, J. M., 112
Bakhtin, M. M., 23
Balbani, A. P. S., 375
Ball, M. J., 59
Ballard, L., 22, 312

¹Note: Page numbers followed by 'n' refer to notes.

- Balogh, J., 272, 303
 Bandi Rao, S., 65
 Baran-Lucarz, M., 83, 87, 88, 90,
 101n7
 Barnard, E., 270
 Barrett, R. P., 177
 Barriuso, T. A., 145
 Barton Laws, M., 393
 Batliner, A., 281
 Baugh, J., 21, 22, 369
 Baxter, G. P., 290
 Baylis, G., 112
 Beattie, G. W., 361
 Beckman, M. E., 383
 Beebe, L., 77
 Beinhoff, B., 33, 127, 134, 142, 176
 Beit-Hallahmi, H., 89
 Benoît, C., 253
 Berkil, G., 211
 Bernstein, J., 272, 303, 321, 322
 Best, C. T., 73
 Bialystok, E., 77–79
 Bigelow, M., 81, 194, 295
 Binnenpoorte, D., 270, 304
 Birdsong, D., 77, 79–81
 Black, E., 346, 364
 Black, R., 257
 Blake, R., 274
 Bley-Vroman, R., 59, 78
 Block, D., 124, 214
 Boers, F., 209
 Boersma, P., 71, 72
 Bolinger, D., 362, 426, 430,
 433–435
 Bongaerts, T., 78
 Bonilha, L., 112
 Boraie, D., 85
 Bornstein, R. F., 48n4
 Bortfeld, H., 243
 Bosker, H. R., 300, 304, 305
 Bosseler, A., 253
 Bosworth, L. A., 88
 Boulaïka, G., 72
 Bourhis, R. Y., 95, 124
 Boves, L., 238, 270, 272, 304
 Bowen, J., 75
 Bowles, A. R., 255
 Boysson, B., 107
 Bradac, J. J., 127
 Bradford, B., 19, 23, 27, 93, 440
 Bradlow, A., 72
 Bradlow, A. R., 199, 200, 414
 Brannon, R., 89
 Brazil, D., 18, 190, 426, 430, 434,
 443
 Bresnahan, M. J., 178
 Brière, E., 75
 Brinton, D. M., 181, 182, 426
 Brooks, G., 385
 Browman, C., 59
 Brown, A., 94, 150, 207, 291, 407,
 414, 426, 438
 Brown, B., 352
 Brown, H. D., 85, 88
 Browne, K., 36, 45, 301, 309, 311,
 425
 Brumfit, C., 37
 Brumfit, C. J., 121, 192
 Brutt-Griffler, J., 345
 Bryant, P., 383
 Bryden, M. P., 61, 100n2
 Burkhalter, C. M., 177, 179
 Bybee, J., 59
 Bygate, M., 293, 296, 331
 Byrd, P., 177–179
 Byrne, J. L., 95

C

- Cameron, D., 18, 359, 360
Campbell-Kibler, K., 20, 26, 27
Canagarajah, A. S., 175
Candlin, C., 358
Cao, R., 211
Carbonaro, M. D., 245
Carey, M. D., 309
Cargile, A. C., 127
Carland, M., 88
Carpenter, R., 251
Carrie, E., 129
Carroll, J. B., 83, 324
Carter, R., 352
Castro-Caldas, A., 81, 383
Catford, J. C., 150, 407
Cauldwell, R., 200, 239, 414, 415,
427, 431, 433, 435, 443
Celce-Murcia, M., 182, 184, 216,
217, 426, 445
Cervone, D., 84
Chalhoub-Deville, M., 296
Chambers, B., 385
Chambers, D., 178, 179
Chang, J., 21
Chapelle, C., 88, 90
Charlton, K., 392
Chartrand, T., 360
Chen, A. H., 324, 352
Chen, L., 324, 352
Cheng, J., 272, 303, 322
Chereau, B. M., 318
Cheung, A., 385
Choi, J-Y., 247
Chomsky, N., 58, 121
Chou, F., 249, 250
Chozick, A., 346, 364
Chun, D., 123
Chun, D. M., 243, 426
Chung, D.-U., 208
Cincarek, T., 281
Cintrón-Valentín, M., 194, 195
Clément, R., 95, 96, 412
Clifford, R., 193
Coetzee-Van Roey, S., 442
Cogo, A., 421
Cohen, M. M., 253
Coleman, J., 433
Collins, 210
Collins, B., 31, 32, 128, 414, 422
Collins, L., 173, 409
Coniam, D., 240, 245
Constantinides, J. C., 177
Cooper, H., 392
Cope, J., 87
Cortes, V., 437
Coste, D., 124
Coulthard, M., 190, 426
Couper, G., 70, 97, 182, 195, 203,
205, 206, 214–217, 291, 406,
407, 418, 445
Coupland, N., 358
Cowie, C., 350, 412
Craik, F. I., 79
Crawford, C., 352
Croft, W., 59
Cruse, D. A., 59
Cruttenden, A., 19, 27, 93, 128,
136, 138–140, 150, 163n2,
301, 411, 414, 422, 430, 433,
434, 439
Crystal, D., 419, 427
Csizér, K., 412
Cucchiari, C., 238, 240, 247, 270,
272, 304, 305
Curnick, L., 456

Curtiss, S., 61, 100n2

Cutler, C., 410

Cutugno, F., 243

D

Da Silva, A. L., 352

Da Silva, R., 268

D'Agostino, P. R., 48n4

Dale, P. S., 81

Dalton, C., 152, 181

Dalton, P., 36, 37, 374, 375, 381

Dalton-Puffer, C., 123, 210

Darcy, I., 409

Dauer, R. M., 142, 435

Davel, M., 270

Davies, A., 42, 298, 332, 333, 351

Davis, S., 385

de Bot, K., 241, 242, 268, 293, 414

de Jong, N. H., 188, 300, 304

de Meo, A., 243, 244

Deci, E. L., 94

DeKeyser, R., 61, 78

Delattre, P., 38

Dénes, M., 270, 272

Dennick, R., 352

DePaulo, B. M., 368

Derwing, T. M., 32, 43–45, 79, 81,
96, 125, 133–135, 148, 150,
181, 186, 188, 201, 218, 238,
245, 273, 300–302, 304, 309,
405, 407, 408, 410–412, 414,
415, 417, 419, 423–425, 431,
441, 443, 449

Deterding, D., 420, 435

Dewaele, J.-M., 85–88

Dewey, M., 421

Di Cristo, A., 429

Díaz-Vera, J., 250

Dickerson, L. J., 77

Dickerson, W., 72

Diepenbroek, L., 408

Dierdorff E., 259

Digalakis, V., 270

Dimroth, C., 193

Dittmann, S., 406

Divi, G., 356

Djkić, M., 88

Dörnyei, Z., 82, 88, 95, 96, 412

Doughty, C., 191, 194, 196

Dovidio, J. F., 21

Drager, K., 27

Dudley-Evans, T., 123

Dull, C., 89

Dunn, J., 107

Dunn, P. K., 309

Dunphy, J., 178, 179

E

Eades, D., 368

Eckert, P., 42, 124, 131

Eckhardt, M., 260

Eckman, F., 76

Eckstein, G. T., 97, 211

Edlund, J., 241

Edwards, J., 123, 383

Eefting, W., 70, 72, 73, 79

Ehri, L. C., 385

Ehrman, M. E., 83

Ehsani, F., 243, 245, 268

Elias, N., 349

Elliot, R., 268

Ellis, N. C., 59, 60, 73, 81, 82, 148,

194, 195, 198, 199, 209,

220n1, 383, 428, 429

Ellis, R., 123, 192, 195, 198, 199,
204, 205, 416
Erard, R., 89
Erb, M., 110, 114, 337
Escudero, P., 71, 72
Eskenazi, M., 243
Esling, J. H., 122, 129, 196
Estebas-Vilaplana, E., 19, 27
Evans, S., 207
Ewert, D., 409

F

Fayer, J. M., 11, 16, 37, 44
Feldman, S., 361
Felps, D., 243, 244
Fernald, A., 66
Field, J., 45, 189
Fietzer, A. W., 85
Fischer, I., 286
Flege, J. E., 70, 72–74, 78–80, 135,
406, 425
Flores, G., 356
Florijn, A. F., 188
Fogassi, L., 98
Foote, J., 408
Foote, J. A., 173, 180, 204, 208,
219, 262, 408–410, 431
Forey, G., 345, 349, 412
Fought, C., 92, 131
Foulkes, P., 27
Fouz-González, J., 220n2, 237, 250,
262, 264, 274, 277n65
Franco, H., 270
Frank, V. M., 255
Fraser, H., 97, 203, 215
Freebairn, L., 379
French, J. P., 367

Freyunik, S., 255
Fridriksson, J., 112
Frieda, E. M., 79
Friginal, E., 349, 351, 430
Fröhlich, M., 85
Frost, D., 225, 456
Fryer, L., 251
Fulcher, G., 36, 45, 290, 301,
307–309, 311–313, 425
Fuleni, I., 107
Furnham, A., 85–87

G

Gallese, V., 98
Gao, X., 290
García, O., 42, 124
Gardner, R. C., 88, 93–95
Garvin, S., 128
Gass, S., 309
Gatbonton, E., 92, 93
Gathercole, S., 383
Gavioli, L., 210
Genesee, F., 122
Gerken, L. A., 59
Gerosa, M., 249
Gheytanchi, A., 98
Gibbon, F., 238
Gibbon, F. E., 374–378, 380, 382
Gilbert, J. B., 126, 144, 182, 184,
237, 352, 426, 431
Giles, H., 92, 95, 124, 127
Gillotti, C., 356
Giuliani, D., 248, 249
Gluszek, A., 21, 134
Godwin-Jones, R., 255, 261
Goh, C. C. M., 189
Goldstein, L., 59

- Golestani, N., 97
Golonka, E. M., 255
Goo, J., 198, 416, 441
Goodenough, D. R., 90
Goodwin, J. M., 181, 182, 426
Gor, K., 188
Gordon, J., 409, 417, 419, 446
Gordon, M., 419
Goronzy, S., 270
Gorsuch, G., 178, 179
Gorsuch, G. J., 178, 179
Goswami, V., 383
Goyen, J. D., 383
Grabe, E., 366, 433, 451n1
Graddol, D., 135, 420
Granena, G., 61, 77, 78, 84, 198,
325, 416
Grant, K., 268
Grant, L., 130, 181, 426
Greatbatch, D., 361, 362
Greaves, W. S., 5, 15, 426, 430
Green, K. P., 73
Greenberg, S., 268
Griffee, D. T., 178
Grimshaw, G. M., 61, 100n2
Griner, B., 181–182, 426
Grodd, W., 114
Guiora, A., 88, 89
Gumperz, J. J., 17–19, 27, 92, 430
Guo, E., 365
Gussenhoven, C., 434
Gut, U., 210
Gutierrez-Osuna, R., 243
- H
- Hacker, C., 249
Hahn, L. D., 187
Hakuta, K., 77, 79
Hall, J., 385
Halliday, M. A. K., 5, 15, 121, 426,
430
Hammerly, H., 145
Han, J., 259, 260, 373
Hanau, C., 10, 159–161, 357, 420,
438
Hansen Edwards, J. G., 72, 81, 91, 92
Hansen, J., 90
Hansen, K., 81, 194, 295
Hansen, L., 90
Harashima, H. D., 255
Hardcastle, W. J., 36, 37, 374, 375,
381
Harding, L., 292, 303, 308, 315,
316, 329, 330, 425
Hardison, D. M., 187, 240, 241,
243, 253, 268
Harley, T. A., 46
Harnsberger, J. D., 368
Harris, J., 257
Harris, R., 42
Harris, Z., 120
Harrison, P., 367
Hawthorne, K., 357
Hay, J., 27
Hazan, V. L., 72
Healy, E. W., 112
Hecht, B. F., 76, 148
Heilman, M., 324
Henderson, A., 173, 174, 176, 180,
201, 218, 408
Heritage, J., 352, 361, 362
Hewings, M., 427, 431, 433
Hieke, A. E., 38, 437
Higgins, D., 324
Higgs, T., 193
Hill, C. J., 245, 372
Hillenbrand, J., 70, 72
Hincks, R., 240, 241, 272, 273, 322
Hinoftis, F., 10

Hirabayashi, K., 145
 Hirata, Y., 243, 268
 Hirose, K., 268
 Hirst, D. J., 429
 Hjalmarsson, A., 254, 255
 Hockett, C., 135
 Hoekje, B. J., 174, 345, 352, 356
 Hoffman, S., 346
 Holland, V., 257, 258
 Hollien, H., 368
 Holtby, A. K., 408
 Holzman, M., 62, 66
 Honey, J., 127, 130
 Honikman, B., 31, 206, 207
 Hood, S., 349
 Hopp, H., 78, 79
 Hornby, A. S., 121
 Horwitz, E. K., 87
 Horwitz, M. B., 87
 Hu, X., 83, 85, 89, 98, 325
 Hulstijn, J. H., 188
 Hutnik, N., 95
 Hyltenstam, K., 295
 Hymes, D., 42, 121
 Hyun, E.-j., 260

I

Isaacs, T., 43, 45, 133, 188, 292,
 298, 302, 308, 311, 312, 318,
 328, 330, 419, 424
 Iverson, P., 63, 406

J

Jackson, S., 188
 Jacobs, E., 352
 Jakobson, R., 407
 Jang, J., 180, 267, 415
 Jang, S., 208

Janicka, K., 411
 Jarosz, A., 408
 Jenkins, J., 41, 43, 123, 128, 131,
 133–136, 138, 140–142, 150,
 162, 176, 181, 189, 298, 299,
 301, 344, 345, 405, 411, 412,
 419–423, 430, 439, 440, 447,
 449
 Jenkins, S., 177
 Jenner, B., 135
 Jensen, F. E., 64
 Jesse, A., 253
 Jo, M., 260
 Johansson, F. A., 75
 John-Baptiste, A., 352
 Johns, C., 190, 426
 Johns, T., 210
 Johnson, J., 78
 Johnson, K., 121, 192
 Johnson, M., 313
 Johnson, R., 188
 Johnstone, B., 124
 Johnstone, M. J., 352
 Jones, R., 345
 Jones, R. H., 207

K

Kachru, B. B., 135, 176, 259, 349,
 419, 421
 Kakehi, K., 200
 Kaltenboeck, G., 240
 Kandel, S., 253
 Kane, M. T., 293
 Kang, B., 260
 Kang, O., 151, 188, 270, 272,
 277n70, 309, 424
 Kanitsaki, O., 352
 Kaplan, J., 257
 Karlner, L. S., 352

- Karp, S., 89
 Karpf, A., 362, 364
 Kassabgy, O., 85
 Kautzsch, A., 225, 456
 Kawai, G., 268
 Kenworthy, J., 126, 133, 422
 Kerswill, P., 27, 68
 Keysar, B., 21
 Kibishi, H., 145
 Kim, D., 260
 Kim, H. K., 269
 Kim, S., 260
 Kim, S-y., 260
 Kirchhoff, K., 269
 Kirkova-Naskova, A., 225, 456
 Klein, R., 79
 Klein, W., 193
 Knodt, E., 243, 245, 268
 Kochanski, G., 433
 Koda, K., 186, 379, 385, 386
 Koehler, K., 188, 304
 Kormos, J., 83, 270, 272, 293, 325
 Körmüçü, S., 191, 192
 Koss, R. G., 356
 Koyama, D., 323
 Krajka, J., 248
 Krashen, S., 121, 122, 192, 373
 Krasinski, E., 11, 16, 37, 44
 Kuhl, P. K., 63, 64, 66, 73, 406
 Kul, M., 411
 Kuo, I.-C., 134, 142
 Kusey, C., 94
 Kusumoto, Y., 208
- L**
- Labov, J., 10, 159–162, 357, 420, 438
 Labov, W., 12, 13, 25–27, 33, 40,
 67, 121, 131, 295, 369
 Ladd, R., 426
 Lado, R., 75, 121
 Lai, E. R., 289
 Laitinen, J., 344
 Lake, C., 385
 Lakoff, R., 359
 Lalwani, A. K., 346
 Lambacher, S. G., 200
 Lambert, S., 208
 Lane, H. L., 88
 Lapkin, S., 263
 Larsen-Freeman, D., 85, 89
 Larson-Hall, J., 61
 Lau, L., 92, 440
 Lave, J., 23
 Laver, J., 31, 206
 Leather, J., 72, 188, 406
 Lee, A., 238
 Lee, J., 180, 188, 204, 267, 415,
 416, 441
 Lenneberg, E. H., 61
 Lennig, M., 272
 Lennon, P., 36, 299
 Leonard, L. B., 375, 376, 378
 Leung, C., 42
 Lev-Ari, S., 21
 Levelt, W. J. M., 46, 190
 Levey, D., 456
 Levinson, S. C., 35
 Levis, J., 173, 183, 196, 209, 238,
 240, 243, 245, 247, 248, 268,
 272, 273, 404, 443, 448
 Levis, J. M., 33, 41, 129, 130, 132,
 133, 145, 163, 181, 187, 292,
 301, 311, 312, 405, 426, 427,
 431, 433, 434, 437
 Lewin, M., 346
 Lewis, B., 379
 Lewis Johnson, W., 258

- Li, K. L., 346
 Li, W., 415
 Li Wei., 88, 124
 Light, J., 253
 Lin, C.-H., 274
 Lindblom, B., 220n2
 Lindquist, H., 210
 Lindsay, J. J., 392
 Lindstromberg, S., 209
 Linell, P., 214
 Lintunen, P., 204
 Liu, S., 74
 Liu, W. Y., 178
 Lively, S. E., 130, 196, 414
 Llurda, E., 133
 Lockwood, J., 345, 349, 350, 412,
 420, 430
 Loeb, J. M., 356
 Logan, J. S., 130, 163n1, 196, 414
 Long, M., 60, 71, 74, 80, 83, 123,
 192, 193, 195, 197, 198, 205,
 416
 Long, M. H., 61, 71, 74, 77, 78, 80,
 84, 85, 89, 188, 192, 194,
 197, 325
 Lord, G., 274
 Loukina, A., 324
 Low, E. L., 156–158, 428, 429, 445
 Lowe, C. T., 20
 Lu, X., 257
 Ly, S., 257
 Lyster, R., 204, 406, 409, 416
- M**
- MacDonald, S., 180
 MacIntyre, P. D., 96, 412
 MacKay, I. R. A., 70, 78, 80, 135,
 406
 Mackey, A., 247
 MacKinnon, G. E., 61, 100n2
 Mai, R., 346
 Mailfert, K., 241, 268, 414
 Major, R. C., 71, 72, 148
 Malone, B. E., 392
 Mannell, R. H., 309
 Marasinghe, C. A., 200
 Marcus, G. F., 65
 Marsh D., 344
 Martens, W. L., 200
 Martin, J. A., 110, 337
 Martin, M., 383
 Martín-Santana, J. D., 346
 Massaro, D. W., 253
 Maye, J., 59
 Maynard, D., 352
 Mayo, S., 393
 Mazlum, F., 421
 McCarthy, M., 209
 McCroskey, J. C., 412
 McDonough, K., 208, 219, 262
 McMahan, A., 377
 McNeilis, K., 356
 Meara, P., 84, 325
 Medina, L., 393
 Mees, I. M., 31, 32, 128, 414, 422,
 426
 Menn, L., 376, 385
 Mey, J., 146
 Meyers, C. M., 178
 Mich, O., 248–250, 268
 Michaud, C., 200, 210, 428
 Miettinen, H., 81
 Milanovic, M., 313
 Miller, B., 79
 Mills, S. V., 204
 Mohammadi, T., 253
 Molholt, G., 200

- Mompean, J. A., 16, 59, 129, 195,
204, 207, 264, 274, 291
- Montemerlo, M., 372
- Montovan, J. C., 375
- Moore, D., 124
- Moran, M., 151, 309
- Morton, S. L., 32, 133, 300, 423
- Moser, D., 97
- Moss, B., 345
- Moussu, L., 133
- Movellan, J. R., 260
- Moyer, A., 22, 33, 80, 410
- Muela-Molina, C., 346
- Muhlenbruck, L., 392
- Mulford, R., 76, 148
- Muller Levis, G., 183, 196, 209
- Munro, M. J., 32, 43–45, 78, 79,
81, 96, 127, 133–135, 148,
150, 181, 186, 188, 200, 201,
238, 240, 245, 268, 273,
300–302, 304, 405–407, 412,
414, 417, 419, 422–425, 431,
441, 443, 449
- Murphy, D., 408, 431
- Murphy, J., 181
- Murty, L., 350
- Mutha, S., 352
- Myford, C., 309
- N**
- Nagano, K., 241
- Naglie, G., 352
- Naiman, N. M., 85, 87, 90
- Nakagawa, S., 145
- Nardi, B., 257
- Nardo, D., 114
- Nebashi, R., 178
- Nelson, C., 44, 133, 300, 422, 423
- Nelson, G. L., 177
- Nemser, W., 75
- Neri, A., 238, 240, 242, 243,
246–250, 269, 270, 274
- Neumeyer, L., 270
- Newport, E., 65, 78
- Niemann, H., 281
- Noels, K. A., 95, 96, 412
- Nolan, F., 366
- Norris, J. M., 197, 198, 416, 441
- Nöth, E., 281
- Novella, M., 198, 416
- Nowacka, M., 411
- Nozawa, T., 79
- Nyquist, J. D., 177
- O**
- Oatley, K., 88
- Ockey, G. J., 323
- O'Connor, J. D., 206, 427
- O'Dell, F., 209
- Odlin, T., 148
- O'Grady, C., 352, 355, 357
- Oh, Y. R., 269
- Ohashi, R., 178
- Okada, A., 208
- Oller, D. K., 63
- Oltman, P., 89
- Origlia, A., 243
- Orion, G. F., 160
- Ortega, L., 124, 197, 198, 416,
441
- Orth, J. L., 178
- Otsuji, E., 124
- Ouni, S., 253
- Oxford, R., 97

Oxford, R. L., 83, 202, 211
 Oyama, S., 78
 Ozawa, K., 241

P

Palmer, H. E., 209
 Papagno, C., 383
 Papansek, M., 107
 Park, S., 260
 Patkowski, M., 78
 Paunović, T., 429
 Pawlak, M., 202
 Peachey, N., 265, 266
 Pearson, P., 268, 272
 Pennington, M. C., 8, 18, 21, 23,
 24, 31, 32, 38, 39, 41, 42, 70,
 73, 91, 92, 101n6, 120, 122,
 124, 129, 130, 146, 153, 156,
 157, 174, 175, 177, 183, 187,
 191, 192, 205, 207, 216–218,
 237–239, 272, 296, 300, 327,
 331, 350, 428, 429, 438–440,
 445, 446, 451n3
 Pennycook, A., 42, 124
 Pervin, L. A., 84
 Pettorino, M., 243
 Phillipson, R., 345
 Pialorsi, F., 177
 Pica, T., 122
 Piccardo, E., 318
 Pickering, L., 178, 188, 240, 243,
 268, 270, 419, 424
 Pierrehumbert, J. B., 426
 Pineau, J., 372
 Pinget, A.-E., 300, 304
 Piske, T., 80
 Pisoni, D. B., 130, 196, 199, 414

Plomin, R., 81
 Plonsky, L., 180, 204, 267, 415
 Pogue, D., 245
 Pollack, I., 253
 Pollack, M., 372
 Ponterotto, J. G., 85
 Porter, D., 128
 Poullisse, N., 293
 Price, H., 349, 412
 Pryor, E., 318
 Purcell, E., 83

Q

Quené, H., 300, 304

R

Rajadurai, J., 442
 Rajagopalan, K., 424
 Rama, P., 257
 Rampton, B., 42, 124, 358
 Ramsay, A., 254
 Ramsey, P. J., 123
 Raskin, E., 89
 Rassaei, E., 90
 Ravid, D., 78
 Reber, R., 48
 Reed, M., 200, 210, 426, 428
 Reinares-Lara, E., 346
 Reis, A., 81, 383
 Reiterer, S. M., 85, 88, 97–99
 Ricento, T., 42
 Richards, D., 134
 Richards, J. C., 76, 121, 130, 148, 175
 Richards, S., 94, 97
 Richardson, D. L., 255
 Richmond, V. P., 412

- Rickford, J. R., 42, 124, 131
 Riggenbach, H., 296, 322
 Rimfield, K., 81, 83
 Rindal, U., 410, 411
 Rittenberg, W., 177
 Rizzolatti, G., 98
 Roach, P., 414, 422, 430, 434–436
 Roberts, C., 88, 90, 345, 352, 354,
 355, 357, 358, 389n3
 Rochet, B., 72
 Rodgers, T. S., 121
 Rodriguez A., 260
 Rogerson, P., 144, 182, 184, 426
 Rogerson-Revell, P., 126, 148, 152,
 181, 183, 185, 202, 206, 213,
 239, 345, 347, 348, 359,
 362–364, 405, 407, 411, 414,
 426, 431, 436, 442
 Ross, C., 178, 179
 Rossiter, M. J., 79, 125, 218, 304
 Rota, G., 114
 Roy, R., 372
 Rubin, D., 178, 188, 210, 424
 Rubin, D. L., 178
 Ryan, E. B., 127
 Ryan, R. M., 94
- S**
- Sabol, M., 257
 Sachdev, I., 48n1, 92, 95, 440
 Sáfár, A., 83, 325
 Saffran, J. R., 65
 Sahota, O., 352
 Saito, K., 45, 75, 77–79, 84, 187,
 204, 298, 325, 406, 409,
 416–419, 424
 Saito, Y., 417–419
 Sanders, T., 300
 Sapon, S., 83, 324
 Sarangi, S., 345, 358
 Sardegna, V. G., 94, 418
 Sassoon, C., 127
 Sato, M., 204, 409, 416, 418
 Savić, M., 429
 Savignon, S., 42, 122, 193
 Sawaki, Y., 323
 Scales, J., 134
 Scharenborg, O., 272, 311
 Scheuer, S., 421
 Schmaltz, S. P., 356
 Schmid, M., 78, 79
 Schmidt, R. W., 77, 85, 97, 195,
 199
 Schoonen, R., 188
 Schultz, T., 269
 Schumann, J., 92, 193
 Schwartz, G., 421
 Schwarz, N., 48
 Scobbie, J. M., 27
 Scott, B., 247, 255
 Scovel, T., 61, 71, 74, 78, 89
 Sebastian, T., 372
 Seedhouse, P., 262, 372
 Segalowitz, N., 36, 92, 299
 Seidlhofer, B., 126, 152, 181, 210,
 345, 411, 440
 Selinker, L., 71, 76
 Setoguchi, E., 323
 Sewell, A., 133, 142, 143, 440
 Shafer, V. L., 69, 70, 72
 Shavelson, R. J., 290, 296
 Shaw, P., 131
 Shaywitz, S. E., 376, 379, 385
 Shea, M., 361
 Shearman, S. M., 178

- Sheldon, A., 72
Sheng, T., 98
Shiki, O., 208
Shin, K., 260
Shockey, L., 68
Sifakis, N. C., 176
Sinclair, J. M., 210
Singh, N. C., 88
Singleton, D., 11, 16
Sinharay, S., 323
Skehan, P., 83–85, 95, 294
Skinner, B. F., 220
Slavin, R. E., 385
Smit, U., 94
Smith, J. A., 177, 179
Smith, K., 177
Smith, K. A., 178
Smith, L., 44, 133, 177, 300, 422, 423
Smith, L. E., 419
Smith, R. M., 178
Soruč, A., 210
Sougari, A. M., 176
Spada, N., 81, 194, 198, 415, 416, 441
St. John, M. J., 123
Stansfield, C., 90
Steidl, S. (Found only in reference)
Steinel, M. P., 188
Sterling, C., 387
Stern, H. H., 85, 325, 368
Stevenson, I., 177
Stockwell, R., 75
Stoel-Gammon, C., 376, 385
Strange, W., 69, 70, 72, 406
Strickland, F., 89
Strik, H., 238, 240, 247, 270, 272, 304, 305
Sumbly, W., 253
Sun, A., 233
Sung, C. C. M., 133
Surface, E., 3, 41, 209, 333, 404
Suter, R., 83
Suzuki, M., 272, 303
Swain, M., 204, 263
Sykes, J. M., 257
Szyszka, M., 94, 97, 211, 212
- T
- Taatgen, N. A., 59, 96
Tamai, K., 208
Tannen, D., 18, 21, 48n3, 366
Tarone, E., 72, 76, 77, 81, 82, 148, 194, 295, 296
Tarone, E. E., 76, 148
Taylor, D. M., 124
Tergujeff, E., 201, 409, 411
Terrell, T., 122, 192
Terrell, T. D., 192
Thelen, E., 61–63
Thompson, T., 356
Thomson, R. I., 79, 96, 180, 181, 199, 200, 239, 304, 412, 414, 415, 418
Thorne, S., 257
Timmis, I., 134
Tipton, S. M., 345, 356
Tohkura, Y., 199, 414
Tollfree, L., 27
Tomalin, B., 15, 349, 420, 430
Tomasello, M., 59, 61, 91
Tomita, Y., 81, 198, 416, 441
Tomlinson, G., 352
Tomokiyo, L. M., 270
Torgerson, C. J., 385

- Treiman, R., 387
 Trofimovich, P., 43, 45, 92, 93, 133,
 173, 188, 292, 298, 302, 308,
 311, 318, 328, 330, 409, 419,
 424
 Trudgill, P., 26, 27, 127
 Truet, T., 107
 Tyler, M. D., 73
- U**
- Underhill, A., 205–207, 407
 Ur, P., 42, 43, 48n5, 126, 132, 135,
 299
 Urzúa, F., 173, 409
- V**
- Vago, R. M., 76, 77, 148
 Van der Zee, K., 85
 van Doremalen, J., 305
 van Es, E., 257
 van Leeuwen, T., 31
 van Lier, L., 313
 Van Moere, A., 272, 303, 322
 van Oudenhoven, J. P., 85
 van Zanten, M., 320, 321, 356,
 389n2
 Vannestål, M., 210
 Velleman, S. I., 62, 63, 65, 66,
 377–379
 Vihman, M. M., 62, 63, 65, 66
 Vijayan, S., 65
 Virnes, M., 260
 Vishton, P. M., 65
 Viswanathan, M., 79
 Vitale, M., 243
 Vo, S., 309
 Vrignaud, N., 253
- W**
- Wachowicz, K. A., 247, 255
 Walker, R., 142, 408, 420
 Walters, A. J., 177, 179
 Wang, X., 200, 240, 268
 Warschauer, M., 257, 274
 Wass, V., 345
 Watanabe, Y., 85, 97
 Watt, D., 27
 Weckwerth, J., 411
 Weinberger, S. H., 77
 Weintraub, M., 270
 Weizenbaum, J., 251
 Wells, J. C., 27, 128, 150, 426,
 440
 Wenger, E., 23
 Wennerstrom, A., 134, 304, 426
 Wennerstrom, A. K., 134, 304,
 426
 Wichmann, A., 427, 431, 434
 Widdowson, H. G., 121, 130, 192
 Wiebe, G. E., 417
 Wiener, H. D., 360
 Wik, P., 254, 255
 Wildner, S., 247
 Wiley, E., 77
 Williams, A., 27
 Williams, J., 191, 194, 196
 Williamson, D., 324
 Wilner, L. K., 10, 15
 Wilson, M., 189
 Windsor Lewis, J., 367
 Winke, P., 22, 292, 309, 312
 Winkelman, P., 48n4
 Winkler, S., 88
 Witkin, H. A., 89, 90
 Wode, H., 76, 78
 Wong, J. W. S., 200, 414, 418
 Wong, W., 194, 196

Wu, S., 134
Wulff, D. H., 177, 198, 209
Wulff, S., 60, 82, 194, 220n1

X

Xi, X., 324

Y

Yates, L., 318
Yau, J. N. W., 31, 350, 359, 420, 430
Yeni-Komshian, G., 74
Yilmaz, E., 210
Yilmaz, Y., 198, 416
Yoon, J. S., 269

Young, R. F., 295, 313, 331
Yu, G., 318

Z

Zampini, M. L., 73
Zarate, G., 124
Zatorre, R. J., 97
Zechner, K., 324
Zielinski, B., 189
Zielinski, E., 318
Zoghbtor, W., 420
Zoghbtor, W. S., 421
Zuckerman, B., 393
Zuengler, J., 23, 77, 92, 150
Zukowski-Faust, J., 177

Subject Index¹

NUMBERS AND SYMBOLS

/s/-fronting, 26

A

Accent, vi, 4, 13, 20–26, 32–34, 38, 43, 45, 68, 70, 71, 78, 80, 82, 84, 92, 94, 99, 124, 126–135, 140, 146, 148–150, 155, 163n1, 176, 178, 204, 206, 207, 210, 213, 214, 239, 243, 244, 246, 247, 270, 271, 274, 289, 297, 298, 301, 304, 307, 309, 312–316, 322, 330, 334n2, 346–350, 355, 357, 359–362, 364–367, 370–373, 388, 389, 401, 406, 411, 415, 419–421, 425, 431, 438, 442, 443, 447

addition, 131, 134
in business contexts, 346–347
familiarity, 45, 269, 307, 309, 330, 425
neutralization, 350
reduction, 131, 134, 289, 350
Accentedness, 32–34, 43, 45, 79, 93, 135, 178, 208, 297–299, 304, 330, 404, 419, 425, 447
Accentuated components, 28
Accentuation, 5, 348, 364, 427, 432, 435, 436
Accuracy, 10, 34–41, 43–45, 72, 83, 84, 88, 93, 122, 125, 132, 152, 160, 182, 186, 188, 193, 194, 241, 245, 248, 249, 255, 262, 263, 268, 270, 273, 274, 289, 292, 294, 295, 297–299,

¹ Note: Page numbers followed by ‘n’ refer to notes.

- 303, 304, 308, 314, 316, 318,
319, 326, 329, 331, 348, 410,
416, 418
- Acoustic analysis, 236, 240, 241,
310, 377, 424
- Acoustic intensity, 5
- Acoustic phonetics, 27, 406
- Acoustic space, 63, 64, 71
- Acoustic trace, 310
- ACTFL Guidelines, 313–314
- Affective factors, 93–96
- Affective filter, 192, 373
- Affiliation, 8, 12, 23, 33, 67, 68, 92,
93, 95, 120, 124, 296, 327,
331, 358, 359, 365
- African, 13
- African American Vernacular English
(AAVE), 21, 369
- Age effects, 75, 77–80
- Age of arrival, 77–80
- Agreeableness, 84, 88
- Alexa*, 237, 263
- Alexia, 374, 380, 382
- Algorithm, 310, 323
- Allophone, 25
- Amalgum English, 136, 138, 139
- American Council on the Teaching
of Foreign Languages
(ACTFL), 312–314
- American English Speech*, 261
- Amplitude (loudness), 28, 246
- Analytical assessment scales, 319
- Animated agent, 257
- Anxiety, 86–87, 94, 192, 373, 381
- Apps, vii, 4, 8, 47, 71, 72, 81, 97,
100, 121, 127, 151, 207,
210, 236–238, 240, 244,
245, 249–252, 254, 257,
259, 261–264, 274, 287,
308, 323, 343–390, 413,
414, 420, 424, 450
- Apraxia, 374, 375, 380, 381
- Aptitude, 82–84, 89, 96, 98, 99,
289, 291, 303, 324–326, 404,
405, 413, 443, 448
- Arabic, 6, 87, 253, 254, 257–259,
321, 434
- Arabic for Global Exchange*, 261
- Arabic talking head, 253
- Articulation rate, 36
- Articulation Station*, 249
- Articulatory flexibility, 88
- Articulatory phonetics, 27
- Articulatory setting, 30–32, 34, 71,
206–208
- Artificial Intelligence (AI), 238, 244,
251, 258, 261, 369
- Artificial intelligence agents, 258
- Artificial palate, 237, 377
- Asian, 13, 19, 178, 420, 429
- ASR-based scoring, 270
- Assessment, vi, vii, 22, 57, 97, 99,
100, 146, 151, 152, 177–178,
235–277, 287–300, 302–313,
318, 319, 321–324, 326–333,
351, 357, 358, 369, 377, 382,
402, 404, 420, 424, 425, 433,
442–444, 449
- Assimilation, 38, 95, 138, 190
- Assistive technology, 262, 369, 370,
372
- Attitudes, 4, 7–9, 15, 22, 30, 48n4,
74, 91, 94–96, 125, 127, 129,
138, 176, 177, 257, 346, 357,
358, 387, 389, 403, 404,
410–413, 421, 422, 425,
427–429, 445, 446, 457
- Audiolingualism, 120–121, 205

- Auditory input therapy, 380
 Auditory learning, 90
 Auditory pattern-finding, 65
 Auditory phonetics, 27
 Auditory traces, 65
 Auditory/perceptual approach, 380
 Australia, 13, 19, 48n5, 368, 384
 Australian, 11, 13, 20, 128, 130,
 213, 239, 298
 Automated scoring, 246, 271, 322
 Automaticity, 10, 31, 39, 41, 48n4,
 86, 151, 218, 297
 Automatic Speech Recognition
 (ASR), 236, 237, 239,
 244–249, 251, 263, 269–273,
 292, 310, 311, 323
 Auto-pilot, 29, 35, 40, 70, 71, 151
 Avatars, 256, 313, 370
 Aviation English, 321
 Awareness-raising, 196, 197, 211
- B**
- Badr*, 253
 Bangladeshi, 92, 93
 BBC pronunciation, 128
 Behaviorism, 120, 121
BetterAccentTutor, 241
 “Big Five” personality dimensions,
 84, 86
 Bilingual, 20, 64, 72, 73, 79, 120,
 122, 123, 130
 Bilingual education, 122, 123, 162,
 349
 Bilingualism, 69, 92, 123, 448
 Black English, 21
 Blocking, 194, 373
 Boston, 12
 Bot, 252
- Bottom-up *vs.* top-down, 187–192
 Brain imaging, 97–99, 382
 Breakdown fluency, 300
 Bridging activities, 209–210
 British, 5, 17–20, 26, 27, 32,
 127–130, 137, 140, 141, 156,
 157, 181, 213, 239, 241, 298,
 325, 361, 411, 428, 433
- C**
- California, 122
 Call centers, vi, 15, 132, 153, 332,
 333, 343, 345, 348–351, 359,
 360, 371, 388, 420, 430, 444
 Cambridge English: Advanced
 (CAE) Exam, 315, 319–321
*Cambridge English Pronouncing
 Dictionary, The*, 237
Cambridge Words, 264
 Canonical babbling, 62, 63
 Canonical syllables, 63
 Cantonese, 6, 16, 17, 150, 434, 435
 CAPT for young learners, 248–250
 Careful speech, 13
 style, 13, 25
 Caretaker talk, 66
 Caribbean, 13, 150, 429
 Case studies, 120, 156–161, 428,
 429, 438, 446
 Casual speech, 13, 26
 style, 13, 25
 Category boundaries, 203, 215, 406,
 445
 Charleston, 12
 Chatbot, 251–252
 Chats, 252, 361
 Chatterbot, 251
 Children’s spelling, 387

- Chinese, 92, 93, 96, 140, 157, 159, 161, 192, 241, 243, 261, 321, 371, 386, 412
- Chinese characters/*hanji*, 386
- Chinese for Beginners MOOC*, 261
- C is for Cow*, 249
- Clarity, 1, 6, 10, 13, 44, 152, 160, 193, 294, 301, 305, 317, 332, 351, 405, 441–442
- Classroom assessment, 153, 302, 326–327
- Classroom-based pronunciation studies, 409
- Classroom-based research, 409–410, 414, 442, 449
- Clear Speech*, 237, 251, 426
- Cleft palate, 238, 375, 376
- Cloze tests, 302
- Cluttering, 374, 375, 380, 381
- Coarticulation, 20, 35, 38, 187, 300, 381, 426
- COBUILD (Collins Birmingham University International Language Database), 210
- Cognitive Aging Hypothesis (CAH), 78, 80
- Cognitive comparison process, 194
- Cognitive errors, 387
- Cognitive fluency, 36, 299
- Cognitive phonology, 205, 214, 291, 445
- Cognitive processing, 59–61, 97, 100n1, 271, 296, 374, 378, 404, 405
- Cognitive strategies, 202
- Cognitive styles, 82, 84, 85, 88, 90
- Coherence, 10, 36, 40, 318, 319, 329, 330
- Commercial off-the-shelf (COTS), 257
- Common European Framework of Reference (CEFR) for Languages, 123, 124, 314–319, 322
- Communicative competence, 6, 39, 42, 43, 47, 96, 121, 124, 125, 143, 177, 193, 312, 326, 331, 350, 356, 421, 447
- Communicative function, 15, 28, 46, 183, 414, 429
- Communicative language teaching (CLT), 121–122, 125, 192, 193, 216
- Communicative pragmatics, 43, 411
- Communicative utility, 65, 151
- Comparative curriculum studies, 446
- Competence, 6, 8, 9, 13, 14, 26, 28, 29, 34, 39, 41–43, 47, 87, 96, 121, 122, 124, 125, 131, 132, 134, 135, 142–145, 155, 162, 174–177, 192, 193, 213, 244, 269, 287–289, 295, 299, 303, 305, 306, 312, 316, 321, 323, 326, 330–333, 346, 348, 350, 356, 365, 386, 389n3, 412, 414, 421, 423, 424, 443, 447
- Componential forms of validation, 293
- Comprehensibility, 43–47, 93, 133, 150, 160, 161, 182, 188, 193, 208, 297, 299, 301–302, 326, 346, 350, 417–419, 423, 424, 431, 447
- Comprehensible input, 179, 192
- Computer-Assisted Language Learning (CALL), 235, 238, 239, 252, 255, 263, 388

- Computer-assisted learning technologies, 123
- Computer-Assisted Pronunciation Training (CAPT), 235, 236, 238–241, 245, 247–250, 252, 254, 255, 260, 262, 268–270, 273, 274, 403, 415
- Computer-based approaches, 380
- Computer-based pronunciation assessment, 269–272
- Computer-generated speech, 236, 325
- Computerized personal assistants, 237
- Connected Speech*, 239
- Conscientiousness, 84
- Consequential validity, 292, 298, 306, 311, 321
- Consonant, vi, 3–6, 10, 12, 14, 25, 28–30, 40, 62, 63, 65, 70, 73, 77, 79, 137–140, 148, 151, 153–155, 160, 161, 182, 184, 201, 203, 207, 238, 240, 272, 297, 298, 374–376, 384, 418, 439, 447
- Construct definition, 290, 308
- Construct-irrelevant variance, 307, 425
- Content and Language Integrated Learning (CLIL), 123
- Contextualization cue, 18, 19, 27, 30, 31, 46, 331, 355, 357
- Contextual variant, 25, 63, 69, 70, 129, 383, 432, 437
- Contrasting L and L learning, 1, 2, 68–69
- Contrastive analysis (CA), 75, 76, 254
- Contrastive Analysis Hypothesis, 75
- Controlled processing, 86, 208
- Conversational Chatbot, 252
- Cool Speech*, 239, 251, 414, 415
- Corpora, 210
- Corpus, 210, 211, 347
- Corrective feedback (CF), 66, 194, 196, 197, 204–205, 216, 247, 249, 327, 328, 416, 417
- Correlation, 78, 87, 88, 90, 101n7, 270, 304, 306, 323, 389, 425
- Cortana, 237, 244
- Council of Europe, 123, 124, 314–317
- Covert contrasts, 297, 377
- Covert prestige, 27
- Creaky voice, 360
- Creole-influenced accents, 150
- Critical listening, 203–204, 208, 211, 215, 406, 407, 445
- Critical period, 61, 78–80, 84, 288
- Critical Period Hypothesis, 78, 80
- Croquelandia*, 257
- Cross-fertilization, 99, 443, 451
- Crossing, 42, 124
- Cue value, 6, 64, 65, 151, 194, 438
- Customer service representatives (CSRs), 15, 332, 349–351, 359
- D**
- Data-driven learning (DDL), 210–211
- Data-driven research, 442–443
- Deaf community, 375
- Deafness, 375, 381
- DEAL*, 254–256
- Deciding teaching priorities, 143–144, 147, 150, 151, 403

- Declarative knowledge, 198
- Decoding, 29, 41, 46, 186, 379,
386, 387, 432
errors, 305
- Delivery, 190, 237, 300, 303, 304,
310, 322, 323, 329, 350–352,
354, 359, 361, 364, 366
- Descriptor, 124, 305, 307, 308, 312,
314, 316–319, 329
- Developmental errors, 76
- Developmental phonological
disorders, 376, 380, 382
- Diagnosis of pronunciation
problems, 327
- Dialogue-based systems, 237
- Dictation, 159, 160, 182, 262, 263,
302, 357
- Digitized model of performance, 310
- Diphthongs, 438–440, 451n3
- Discourse intonation, 179, 190
- Discrete-item forms of assessment, 291
- Discrimination activities, 206
- Disfluency markers, 36
- Dissemination, research findings,
440–441
- Double-voicing, 124
- Dragon Dictate*, 236, 263, 371
- Dragon Naturally Speaking*, 236, 245,
263, 371, 372
- Drilling, 121, 200, 238
- Drug therapies, 381
- Duolingo*, 240, 251, 252
- Duration (rhythm), 4, 28
- Dutch, 247, 270, 300, 304, 305,
321, 442
- Dysarthria, 374, 380
- Dyslexia, 369, 376, 379, 380, 382
- Dyspraxia, 374, 375, 380
- E
- Ear witnesses, 366
- Edmodo*, 262
- Education levels, 7, 81
- Educational factors, 75, 81–82
- Educational service robots, 373
- Effectiveness of CAPT, 267–269,
274, 415
- Electropalatography (EPG), 237,
377, 380
- ELIZA*, 251
- ELSA Speak*, 251
- Embodied Conversational Agents
(ECAs), 252–255
- Empathy, 17, 85, 88–89, 96–98,
332, 350, 355, 357
- Employment, vi, 21, 24, 43, 127,
288, 299, 306, 312, 333
- English Accent Coach*, 239, 415
- English as a lingua franca (ELF), 95,
123, 126, 132–136, 138,
141–143, 162, 176, 181, 210,
345, 388, 402, 403, 411, 412,
419, 421–423, 429, 439, 447
- English as an additional language,
136, 163n2
- English as an international language
(EIL), 42, 44, 95, 123, 130,
136, 138, 141, 155, 158, 176,
183, 189, 298, 299, 301, 388,
419–423, 440, 444, 447
- English education robots, 260
- English File Pronunciation*, 250, 251
- English for Kids*, 248
- English for Medicine, 264
- English for Specific Purposes (ESP),
123, 125, 153, 159, 160, 264,
305, 402

- English language teaching (ELT)
policy, 420
- English Pronunciation Teaching in
Europe Survey, 408
- Entrenched/entrenchment, 63, 67,
68, 70, 71, 73, 88, 92, 139,
194, 421
- Epenthetic vowel, 203
- Equivalence classification, 70, 76,
82, 88, 447
- Error tolerance, 149–150
- Ethnic group affiliation, 93
- Ethnography, 446
- Europe, 123, 124, 314, 316, 344
Evernote, 263
- Expanding circle, 135, 156–158,
259, 421
- Expansion stage, 62
- Experimental/quasi-experimental
design, 409
- Explicit FFI, 195–200
- Explicit learning, 59, 60, 81, 96, 99,
190, 194, 199, 220n1, 447
- Extensive listening, 190, 447
- Extravert/extraversion, 84–86, 96
- Extrinsic motivation, 94
- Eye-tracking methodology, 330
- F**
- Facebook*, 237, 264
- Falling tone, 9, 17–19, 179, 363,
430, 431, 433, 434, 449
- Field independence (FI), 85, 89, 90
- Filipino, 15, 20, 21, 350, 351, 359,
430
- Finland, 408
- First *vs.* Second language acquisition,
58–61
- Fluency, vi, 34–41, 44, 45, 47, 48n4,
85, 86, 89, 93, 133, 134, 145,
147, 151, 154, 179, 182, 186,
188, 189, 201, 208, 209, 240,
257, 262, 272, 289, 294, 295,
297, 299, 300, 303–305, 308,
309, 314, 316, 318, 319, 322,
323, 326, 329, 330, 347, 382,
416, 418, 431, 441, 442
- Focus on form, 81, 82, 90, 99, 120,
122, 123, 125, 158, 188, 194,
197, 198, 208, 216, 295, 416,
445
- Force of articulation, 5
- Forced output, 263
- Foreign accent, 21, 34, 70, 71, 78,
80, 82, 84, 127, 134, 148,
243, 315, 316, 320, 330, 347,
389, 406
- Foreign language classroom anxiety
(FLCA), 87, 96, 101n7
- Forensic linguistics, vii, 344,
366–369
- Forensic phonetics, 367, 368
- Formative assessment, 327, 329,
334n1
- Form-focused instruction (FFI), 76,
90, 151, 194–197, 199, 200,
204, 205, 208, 219, 220n1,
328, 402, 403, 416–418, 441,
449
- Forvo*, 264, 265
- Fossilization, 71, 73, 82, 249
- France, 408
- French, 14, 16, 20, 29, 32, 86, 88,
127, 149, 241, 252, 265, 321,
348, 367, 372, 430, 434, 435,
442
- French Digital Kitchen, The*, 262, 372

- Functionality, 65, 194, 249, 254,
 263, 273, 299, 371, 438
 Functional load, 150–151, 154, 247,
 403, 407, 438
 Functional magnetic resonance
 imaging (fMRI), 97, 443
 Functionally Native Proficiency, 314
 Fundamental difference hypothesis,
 59
 Fundamental frequency contours,
 268
 Future research, vi, 220, 332, 402,
 445–448
- G**
- Gaming, 255–259
 Gap
 between research and practice,
 219, 405, 406
 current knowledge base, 440–445
 General American (GA), 126–129,
 132, 183, 246, 429, 439
 General British English (GB), 19,
 126, 128, 132, 156, 183, 246,
 360, 429, 438–440
 German, 14, 19, 20, 33, 89, 97–99,
 127, 206, 211, 252, 371, 434
 Germany, 408
 Gilbert, J.B., 126, 144, 182, 184,
 237, 352, 426, 431
 Given/new, 157, 158
 Global English identity, 95, 132
 Glottal stop, 12, 20, 26, 440
 Goals, vi, 8, 10, 28, 34, 38–42, 44,
 47, 69, 91, 93, 95, 97, 119,
 121–136, 139, 141, 143, 146,
 151, 153–155, 159–162, 176,
 186, 195, 197, 201, 213, 215,
 220n1, 239, 244, 248, 251,
 257, 274, 297, 304, 311, 326,
 331, 350, 374, 402–404, 415,
 420, 421, 425, 440, 441, 447,
 450
 Good pronunciation learners, 94,
 212, 448
 Good pronunciation user, 212
Google Assistant, 263
Google Drive, 262
Google Groups, 261
Google Hangouts, 261
Google Translate, 263
 Gooing stage, 62
 Grapheme, 29, 185, 263, 324, 383,
 386
 Grouping, 15, 28, 147, 179, 262,
 300, 329, 330
- H**
- Habit, 120, 121, 205–207
 Hawaii, 21
 Hawking, Stephen (Professor), 369,
 370, 373
 Health communication, 344,
 352–358
 Hearing loss, 374, 375
 Hesitators, 36, 45, 300
 High ability language imitators, 88
 High-involvement conversational
 style, 366
 High rising tone (HRT), 19, 92,
 440
 High-stakes tests, 180, 270, 442
 High variability perceptual (or
 phonetic) training (HVPT),
 130, 163n1, 196, 199, 200,
 208, 239, 414, 415, 418, 449

- Homonym (homophone), 30
- Hong Kong, v, 16, 17, 20
- Human-machine interaction, 253
- Human rater, 270, 289, 292, 297,
300, 306–309, 316, 320, 322,
442
- I
- iBT
- Practice Online test, 188, 272,
323
- Speaking section, 322, 323
- Identity, v, vi, 2, 7, 8, 23–25, 27, 30,
42, 43, 47, 67–68, 91–93, 95,
99, 100, 101n6, 120, 123,
124, 128, 131, 132, 134, 142,
143, 146, 157, 174–177, 182,
201, 214, 243, 296, 327, 331,
343, 344, 350, 358, 359, 388,
401, 403, 405, 410, 421, 443,
448
- Image, 8, 23, 67, 175, 207, 253,
259, 348, 358–362, 364, 366,
401
- Imitation, 14, 66, 67, 71, 72, 82, 88,
90, 97, 98, 127, 189, 190,
192, 204, 207, 212, 235, 238,
243, 245, 246, 263, 269, 273,
274, 288, 289, 294, 295, 303,
311, 322, 325, 326, 331, 372,
413
- tasks, 82, 97, 289, 311
- Imitative abilities, 61, 99
- Immersion, 101n4, 121
- Impact, vi, 9, 10, 22, 24, 45, 75, 86,
91, 100, 120, 147, 152, 154,
158, 162, 163, 176, 181, 182,
186, 187, 193, 194, 198, 201,
212, 262, 267, 287, 292, 321,
333, 344–346, 348, 351, 362,
364, 366, 369, 372, 374, 384,
401, 403, 404, 416, 425, 431,
432, 436, 447, 450, 451
- Implicit FFI, 195, 196
- Implicit learning, 59, 60, 68, 69,
129, 192, 199, 200, 204, 205,
447
- Impulsivity, 85
- Inaccuracy, 35, 36, 41
- Indian, 17, 18, 213, 325, 350, 429,
430
- Individual differences, 58, 67–68,
75, 78, 82, 83, 87, 90, 97, 99,
100, 307, 332, 405, 412
- Individualized instruction, 100, 448
- Induction, 65
- Infant-directed speech, 66
- Innateness, 121
- Inner-circle, 176
- Input, 59–65, 67, 68, 74, 75, 80–82,
84, 85, 90, 92, 93, 99, 100,
100n2, 101n5, 129, 139, 186,
190–193, 195–199, 202, 204,
208, 211, 217, 220n2, 237,
241, 243–245, 249, 251, 252,
257, 261, 262, 264, 272, 274,
292, 324, 373, 414, 416, 417,
424, 445–447
- enhancement, 100, 196, 220n2
- form, 193
- modules, 60
- Input-flooding, 191, 196
- Input-heavy listening approach, 191
- Instance learning, 60, 61
- Instructional innovation, 409
- Instrumental motivation, 94, 95
- Intake, 28, 198

- Integrative motivation, 94, 95, 412
 Integrativeness, 88, 94, 95, 412
 Intelligence, 26, 34, 40, 89, 206,
 238, 363, 368
 “Intelligent” language tutors, 251
 Intelligibility, vi, 6, 21, 38, 43–45,
 47, 71, 100, 126, 131–143,
 145, 147, 149, 154, 162, 177,
 182, 186, 189, 193, 197, 201,
 243, 247, 253, 255, 269, 289,
 294, 295, 297, 299–304, 308,
 309, 314, 317–319, 323, 326,
 327, 329–332, 343, 345,
 347–349, 351, 357, 370, 388,
 389, 389n3, 403, 404, 407,
 411, 415, 419, 421–425,
 430–432, 436, 439, 441–443,
 447
 Intelligibility Principle, 132
 Interaction, 2, 17, 22, 23, 40, 44,
 63, 66, 67, 72, 80, 85, 91, 93,
 96, 101n4, 121, 124, 133,
 152, 188, 192, 193, 203, 204,
 210, 216, 217, 246, 251, 259,
 288, 295, 301, 303, 311, 312,
 319, 320, 328, 331, 345–347,
 352, 387, 388, 416, 417, 424,
 427, 434, 442–444, 449–451
 hypothesis, 197, 204
 Interactional adjustments, 197
 Interactional competence, 295, 323,
 331
 Interactional dynamics, 350, 432
 Interaction-based research, 348
 Interagency Language Roundtable
 (ILR) Scale, 313, 314
 Interference, 75, 76, 81, 92, 305,
 437
 Interlanguage, 71, 72, 76, 309
 International business meetings,
 347–348, 442
 International English, 42, 136, 138,
 140, 150, 439
 International English Language
 Testing System (IELTS), 302,
 315, 318, 319, 322, 329, 331,
 442
 Speaking exam, 318, 319, 329
 International identity, 95
 International medical graduates
 (IMGs), 10, 156, 159–161,
 312, 355, 357
 International Phonetic Alphabet
 (IPA), 161, 204, 215, 324
 International teaching assistants
 (ITAs), 174, 177–179, 312,
 418
 Internet of Things, 372
 Interpersonal meaning, 182, 350,
 426, 430
 Interpretability, 46, 47, 423, 431
 Inter-rater reliability, 307, 330
 Intervention studies, 409, 419, 449
 Intonation, vi, 4–6, 9, 15, 17–20, 23,
 28, 31, 33, 36, 66, 83, 138,
 140, 147, 149, 153, 157, 160,
 178, 182, 183, 187, 188, 190,
 200, 202, 204, 210, 240, 241,
 243, 267, 268, 297, 302, 303,
 315, 317, 319, 323, 349–352,
 354, 355, 357, 359, 360, 408,
 414, 425–434, 440, 447, 449
 Intra-rater reliability, 306, 425
 Intrinsic motivation, 94
 Introvert, 85–87
 iPhone, 252, 371
IROBI, 260
 Isochronous rhythm, 435

- J**
Jabberwacky, 251
 Japan, 259
Journal of Second Language Pronunciation, 219, 448, 451n4
- K**
 Kachru, B. B., 135, 176, 259, 349, 419, 421
 Key research themes, 402–405
 Kinaesthetic, 205–207, 380
KjennMEG (KnowME), 369
 Korea, 259
- L**
 L1-based motor routines, 70
 L1 literacy, vii, 84, 344, 383
 L1 transfer, 70, 71, 75–77, 139, 148, 185, 186, 191, 203, 415, 447
 L2 environment, 80, 101n4
 Labov's (1966) New York City department store study, 12–13, 40, 295
 Language
 acquisition processes, 76, 121–122
 aptitude, 83, 84, 291, 324, 404, 443, 448
 ego, 89
 laboratory, 121, 122
 proficiency scale, 314–318
 teachers who are L speakers, 2, 174–177
 universals, 76, 139, 148
 Language Acquisition Device (LAD), 58
 Language Analysis for the Determination of Origin (LADO), 368
 Language-like behavior, 295
 Latin American Spanish, 23
 Learnable, 60, 405, 429
Learn English Kids, 248
 Learner errors, 77, 147–148
Learn German Now, 247
 Learning
 processes, 60, 64, 68, 69, 73, 75–77, 81, 94, 97, 148, 186, 190, 199, 205
 skills, 59, 81–85, 89, 97, 403
 strategies, 69, 82, 96, 97, 100, 121, 174, 202, 203, 211, 212
 styles, 82
 styles inventory, 202, 211
 Length of residence, 80, 101n4
 Lexical chunks, 39, 209
 Lingua franca, 42, 43, 130, 132, 135, 136, 138, 140, 143, 298, 299, 301, 328, 331, 344, 345, 420, 440, 444, 447
 Lingua franca core (LFC), 44, 136–138, 141–143, 150, 420, 421, 439, 447
LinguaGraph, 237, 377
LinguaView, 237, 377
 Linguistic cue, 13, 367
 Linguistic marker, 13
 Linguistic/phonological approach, 380
 Linguistic stereotypes, 20
 Linguistic stereotyping, 20, 34, 150
 Linguistic tolerance, 124
 Linking, 9, 20, 29, 38, 48, 120, 182, 184, 237, 294, 303, 416–418, 426, 437, 440

- Listen-and-repeat, 202, 206, 291
 Literacy, 81, 82, 89, 145, 185–186,
 215, 295, 324, 373, 383, 385,
 386
Livemocha, 274
 Liverpool, 150
 LLAMA D, 84, 325
 LLAMA Language Aptitude Test, 84,
 325
 London, 92, 183, 352, 367, 430
 Long and short vowels, 268
 Longitudinal study, 148, 409, 418,
 441, 446
 Long-term memory, 59–61, 68, 194
 Low phonological awareness, 379
 Lower-middle class, 13
- M**
- Macedonia, 408
 Machine learning, 251, 370
 Machine-ratable measures of L
 speech, 2, 304
 Mancunian accent, 360
 Mandarin, 134, 211, 261, 415, 434
 Markedness Differential Hypothesis
 (MDH), 76
 Massive Open Online Courses
 (MOOCs), 261
 Matched guise technique, 26
 Maturation, 83
 Maturation changes, 60, 62, 67
 Maturation constraints, 77
 Maturation schedule, 61
 Maximal/multiple oppositions, 380
 Meaning/semantics, 33, 40, 82, 426
 Memory strategies, 203
 Message form, 18
 Message function, 18
 Messaging, 1, 264
 Meta-analysis, 267, 415, 418
 Meta-analytical review, 188, 384
 Metacognitive awareness, 200, 203,
 428
 Metacognitive strategies, 202, 203,
 211
 Metamessage, 18, 19, 22, 27, 29–31,
 42, 46, 48n3, 303, 331, 350,
 388, 427, 449
 Metathesis, 377, 378
 Metrolingual identity, 124
 Micro-level listening practice, 438
 Microworld, 257
 Migration, 377, 378
 Minimal pair contrast therapy, 380
 Minimal pairs, 5, 16, 60, 150, 182,
 238, 263, 289, 291, 294, 302,
 326, 380, 408, 436–438, 446,
 449
 Mirror neuron system, 98
 Miscommunication, 10, 11, 17, 22,
 39, 293, 301, 327, 356
 Misunderstanding, 2, 11, 15, 17, 18,
 22, 74, 132, 350, 352, 420,
 430, 437, 442
 Mobile games, 255
 Mobile-assisted language learning
 (MALL), 250
 Mock-standard accents, 124
 Model, 23, 32, 41, 43, 57, 63–65,
 71, 83, 92, 95, 96, 119,
 125–134, 142–144, 146,
 158, 159, 162, 163, 163n1,
 175, 176, 181, 185, 190,
 200, 203, 204, 212, 213,
 215, 241–246, 259, 261,

- 267, 269–272, 289, 292,
297, 298, 310, 312, 316,
321–324, 327, 328, 360,
367, 369, 402, 403, 410,
414, 415, 417, 420, 421,
439, 447
- Model accent, 129
- Modern Language Aptitude Test
(MLAT), 83, 84, 89, 324, 325,
332
- Mondly*, 252
- Monitor Model, 192
- Morphemes, 5, 7, 302
- Mother tongue maintenance, 122
- Motivation, 8, 74, 82–85, 88,
93–97, 99, 100, 131, 147,
153, 186, 201, 217, 249, 252,
255, 260, 410, 412, 413, 428
- Motor/articulatory approach, 380
- Motor plan, 64
- Motor template, 31, 35, 40, 67, 71,
72
- Multicultural Personality
Questionnaire, 87
- Multilingualism, 42, 69, 123, 162,
213–214, 248
- Multilingualism–multiculturalism, 123
- Multilingual talking head, 253
- Multiphonology, 42, 124
- Multiplayer online game (MMOG),
255, 257
- Multiple measures of pronunciation,
303, 305
- Musical ability, 83
- Myths, 128, 403, 431–440
- Myths and misconceptions,
pronunciation teaching and
learning, 431–440
- N
- Native Language Magnet (NLM)
model, 63
- Nativeness, 41–43, 132, 331
- Nativeness Principle, 132
- Natural approach, 122, 192
- Natural language processing (NLP),
244, 258, 323, 369
- Natural phonological processes, 76
- Needs analysis, 123, 144–146
- Negative evidence, 66
- Negative transfer, 75
- Negotiation for meaning, 197, 416
- Neurolinguistic programming, 67
- Neuroticism, 84, 86–87
- Newborns, 62
- New York City, 12, 13, 33, 40, 295,
364
- New Zealand, 19, 48n5
- Next-generation speaking assessment
by machine, 322–324
- Nigerian, 130, 354
- Non-regional pronunciation, 32,
128
- Non-rhotic, 12, 13, 33, 137, 141,
213
- Nonsense word repetition tasks, 89
- Nonstandard pronunciation, 11,
16
- Non-verbal cues, 358
- Norms, 10, 34, 38, 75, 134, 140,
141, 143, 146, 147, 176, 263,
297, 299, 311, 334n1, 421,
422, 428
- North America, 19, 23, 141, 209,
411, 439
- North German, 19
- Northern England, 13, 439

O

Online intelligent tutor, 252
 Onset, 384
 Openness, 84, 88, 89, 94, 95, 434
 Operationalization, 290
 OPI exam, 312
 Oral Proficiency Interview (OPI),
 313–314
 Orthographic system, 29, 331, 386
 Output form, 192, 193
 Output hypothesis, 204
 Output modules, 60
 Outsourced call centers, 332, 351
Oxford Words, 264

P

Pakistani, 17, 430
PaperPort, 263
 Paralanguage, 30–32, 426
 Pause fillers, 36, 37
 Pedagogy-driven materials, 238
 Perceived fluency, 36, 299, 304
 Perception, vi, 2, 9, 14, 19, 24–29,
 36, 37, 40–41, 43, 45, 58,
 60–65, 67–74, 77, 88, 90, 91,
 98, 101n3, 133, 143, 144,
 146, 178, 182, 184, 187, 189,
 193, 194, 196, 197, 199, 203,
 208, 214, 215, 240, 243, 251,
 254, 262, 268, 288, 289, 291,
 297, 299, 300, 302–303, 311,
 312, 324–326, 328, 330–332,
 348, 355, 357, 377, 380, 406,
 410, 416, 421, 423, 438
 Perceptual accuracy, 41, 45, 267
 Perceptual automaticity, 41
 Perceptual fluency, 41, 45, 48n4

Perceptual intelligibility, 302
 Perceptual magnet effect, 63, 69,
 406, 407
 Perceptual prominence, 5
 Perceptual training, 100, 199, 208,
 267, 414, 418
 Perceptual warping, 64
 Performance-based assessment, 289
 Personality, 30, 82, 84–89, 91, 96,
 97, 99, 344, 358, 404, 412,
 413, 448
 Personal robots (PR), 259, 360
 Peruvian, 159, 161
 Phonation stage, 62
 Phone, 76, 82, 255, 263, 266, 321,
 322, 345, 351, 368, 371, 384,
 412
PhonePass, 272, 273, 321, 322
 Phonetic coding ability, 83, 89
 Phonetic script, 324, 325
 Phonetic variants, 25–27, 29, 46
Phonetizer, 369
 Phonics, 249, 380, 384, 385, 432
Phonics Studio, 249
 Phonological competence, 8, 132,
 424
 Phonological decoding, 379, 386
 Phonological fluency, 38–40, 151,
 188, 300, 347
 Phonological impairment, 289, 344,
 374–376, 378, 382
 Phonological loop, 60, 98, 383
 Phonological process tests, 377
 Phonological working memory, 60,
 89
Phonology and Phonetics Review, 239
 Phonology in L learning, 1, 61–68
 Physicality, 205–207

- Plurilingualism, 42, 124, 213–214
 Pluriphonology, 42, 124
 Poland, 408, 411
 Politeness, vi, 16, 19, 27, 204, 359
 Politically speaking, 361–366
 Positive transfer, 75
 Post-input, 196, 197
 Postvocalic /r/, 12, 13, 27, 40, 350
Praat, 236
 Pragmatic competencies, 323
 Pragmatic effects, 8
 Pragmatic force, 46, 157, 193
 Pragmatic meaning, vi, 5, 15, 29, 31–32, 46, 182, 348, 359, 426–428
 Pragmatics, 190, 327, 389n3
 Pre-input, 196
 Prelinguistic vocalizations, 62
 Prestige variant, 26
 Prioritizing pronunciation teaching areas, 147–155
 “Proclaiming” intonation, 18
 Productive competence, 42
 Proficiency test, 291, 302, 305–306, 313–314
 Prominence, 5, 28, 179, 187, 190, 202, 209, 362, 363, 427
 Pronunciation anxiety (PA), 87, 96, 101n7
Pronunciation for Communication MOOC, 261
 Pronunciation Learning Strategies Inventory, 202, 211
 Pronunciation learning strategies of high performance learners, 211–212
Pronunciation Power, 237, 240–242, 245, 246
 Pronunciation pragmatics, 74, 330–331
 Pronunciation/speech imitation talent, 97
 Proprioception, 205–206, 403, 407
 Prototype, 63–65, 406
 Pruning, 64
 Psycholinguistics, 36, 99
- Q**
 Québec, 180, 409
 Québec Francophones, 93
- R**
 Racial profiling, 21
 Rapport, 178, 179, 182, 320, 331, 332, 351, 359, 366, 427
 Rating scale development, 290, 308
 Rating scales, 290, 306, 308–309, 329, 330, 443
 Reading, 77, 81, 89, 145, 148, 160, 179, 185, 186, 191, 201, 208, 219, 252, 253, 260, 261, 267, 272, 291, 294, 296, 305, 321, 322, 369, 373, 376, 378, 379, 382–387, 408, 418, 424
 Received pronunciation (RP), 126–129, 139, 362
 Receptive, 42, 134, 144, 192, 414
 Redundancy, 10, 194
 “Referring” intonation, 5, 18
 Reflectivity, 85, 88
 Regional accent, 25, 130, 131, 155
 Regional variant, 25

- Reliability, 273, 291–295, 297, 298,
305–311, 322, 323, 330, 333,
371, 425, 444
- Repair fluency, 300
- Repetition, 37, 66, 71, 121, 125,
179, 185, 196, 202, 207, 212,
235, 238, 300, 303, 311, 315,
317, 326, 366, 374
- Rhotic, 12, 13, 33, 40, 137, 141,
213
- Rhythm, 4–6, 28, 33, 63, 83, 138,
187, 190, 201, 208, 240, 241,
266, 317, 323, 362, 416, 417,
434–436
- Rime, 384
- Rising tone, 9, 17–19, 92, 155, 363,
427, 433, 434, 440, 449
- Risk-taking, 85, 122
- Robot-assisted language learning
(RALL), 259–260
- Roleplays, 190, 248, 256, 260, 274,
355
- Rosetta Stone*, 274
- S**
- Saliency, 194, 380
- Salient, 33, 60, 64, 66, 70, 82, 153,
175, 191, 217, 308, 379
- Santa Barbara Corpus of Spoken
American English, 210
- Sat navs*, 371
- Savannah, 12
- Schwa [ə], 38
- Scotland, 13, 439
- Scouse accent, 150
- Second language acquisition (SLA),
4, 47, 58–61, 88, 89, 96,
100n1, 121, 122, 156, 162,
192–195, 198, 209, 214, 370,
406, 448
- Segmental errors, 11, 189, 442, 447,
449
- Self-esteem, 85, 376, 382
- Sensitive periods, 61, 77, 78
- Sentence types, 427, 433–434
- Shadowing, 204, 208–209, 219,
220n3, 262, 267
- Short-term memory, 60, 85, 86
- Signing, 375
- Simulations, 177, 255–259, 274,
321
- Singaporean, 130, 156, 157
- Siri*, 237, 244, 263, 266, 371
- Skype*, 264
- SLA theory, 121, 122, 156, 165,
192, 214
- Smartphones, 217, 244, 250, 251,
255, 262, 263, 371
- Smart technologies, 261, 370, 372
- Smiling voice, 360
- Social attribute, 27
- Social distance, 101n5
- Social identity effect, 346
- Social learning theories, 214–217
- Social meaning, 2, 15, 27, 29, 43,
46, 74, 147, 326, 389n3, 446,
447
- Social media, 130, 237, 274, 370,
388
- Social skill, 359
- Social status, 12, 20, 24, 128, 367
- Social strategies, 203, 211
- Social variant, 25, 26, 33, 139
- Socioeconomic status, 13, 27
- Sociophonetics, 13, 26, 27

- Solidarity, 27, 101n5
 Sound-focused listening, 186
 Sound recognition, 325
Sounds, 237, 240, 251
Sounds of Speech, 239
 Sound-spelling correspondence, 29,
 77, 82, 89, 183, 185, 186,
 203, 289, 383–385
 Sound waves, 27, 28, 291
 Southeast England, 13
 South Korea, 260
 Spain, 87, 129, 264, 347, 408, 411
 Spanish, 14, 19, 20, 23, 29, 129, 137,
 147, 148, 174, 247, 252, 257,
 321, 347, 408, 434, 435, 442
 Speaker
 agency, 42, 120, 123–124
 authentication, 367
 identification, 366, 367
 profiling, 367
 verification, 367
 Speaking style, 68, 86, 101n5, 271,
 327, 364
Speak It, 101n2, 370
 Spectrogram, 236, 242, 243, 367
 Speech
 analysis, 236, 246, 268, 292, 368
 corpora, 210
 melody, 5, 426
 perception, 27, 29, 40, 41, 60,
 62, 63, 65, 67, 69, 98, 189,
 303, 324, 416
 processing effect, 346
 production, 28, 29, 31, 34–40,
 46, 60, 68–70, 86, 293, 299,
 369, 374, 377, 381, 382
 synthesis, 236, 237, 252, 262,
 369, 370, 372, 373
 technologies, 122, 236–238, 244,
 251, 344, 369–373
 therapy, vii, 235, 237, 238
 training, 237, 346
 Speech accent
 accent archive, 210
 advertisements, 359
Speechlogger, 263
SpeechRater, 272, 273, 323, 324
 Speech-to-text (STT), 236
 Speed fluency, 300
 Spelling, 29–30, 140, 204, 249, 324,
 325, 373, 376, 378, 383–385,
 387
 clues, 324
 pronunciation, 145, 148, 185
 Spoken Error Tracking System
 (S.E.T.S.), 246, 247
 Spontaneous speech, 13, 72, 144,
 145, 270, 296, 304, 305, 323,
 379
 Stance, vi, 236, 358, 424
 Standard English, 22, 128
 Standardized speaking tests, 312
 Standardized test, vii, 90, 179,
 288, 291–293, 312, 325,
 328, 333
 Standard language, 32, 91, 121, 298,
 432
 Standard Southern British English
 (SSBE), 128
 Step 2 clinical skills (CS)
 performance test, The,
 320–321, 389n2
 Stereotyping, 19–22, 34, 150, 346,
 410, 421
 effect, 346
 Strategic competence, 8

- Stress, vi, 4–6, 15, 16, 28, 33, 37,
38, 57, 76, 82, 86, 137, 138,
147, 149, 152, 154–158, 160,
180, 182, 184, 187–190, 196,
197, 200–202, 207, 211, 212,
240, 241, 257, 265–267, 287,
289, 297, 300, 302, 303, 305,
315, 317, 319, 343, 348, 349,
352, 354, 355, 362–365, 367,
368, 375, 379, 380, 384, 401,
407, 416, 417, 420, 425–430,
432, 434–436, 446, 447
 on numbers, 15
- Stress-timing, 240, 435, 436
- Structuralism, 120
- Structural language teaching, 120–121
- Structured listening, 380
- Style-shifting, 42, 77, 124, 214, 331
- Styling, vii, 358, 359, 366
- Supiki*, 251
- Swedish, 254
- Switzerland, 408
- Syllables, 3, 5, 7, 16, 30, 35, 37, 38,
62, 63, 65, 140, 152, 184,
185, 187, 201, 215, 255, 265,
300, 325, 326, 363, 374, 375,
378–381, 384, 429, 434–436
- Syllable-timing, 381
- Sync*, 371
- Synthesis-by-rule, 237
- T**
- Tactical Iraqi*, 258, 259
- Taiwan, 259
- Talking Heads, 252–255
- Talk to Me*, 370
- TalkTyper, 263
- Task
 complexity, 293, 294
 constraints, 77
 difficulty, 293, 294, 326
- Task-based Language Teaching
 (TBLT), 123, 125, 192, 197
- Task-oriented input, 196
- Task-sampling variability, 296
- Teachability/learnability, 152–153
- Teacher identity, 174–177
- Teacher surveys, 408–410
- Teaching and learning activities and
 techniques, 201–203
- Teaching effectiveness research, 188,
267, 404, 416
- Teaching sequence, 216, 217, 445
- Technologies, vii, 100, 122, 123,
125, 156, 211, 217–218,
235–277, 321, 323, 344,
368–373, 377, 382, 402,
413–415, 448
- Technology-driven materials, 238
- Tell Me More/Talk to Me Kids*, 249
- Templates, 35, 60, 64, 65, 67, 70–73
- Temporal fluency, 36, 151, 299, 300,
304
- Test of English as a Foreign
 Language (TOEFL), 188, 272,
273, 302, 303, 315, 322–324,
442
 Speaking scale, 303
- Test validity, 291
- Textbooks, v, 127, 144, 180, 183,
190, 201, 207, 408, 413–414,
422, 426, 433–435, 440, 441,
450
- Theory of mind, 98
- Three-part lists, 152, 362

Threshold level, 10, 34, 175, 177, 423
 Timing, 36, 272, 299, 300, 310, 329, 375, 377, 381, 435
Tokbox, 264, 265
 Tolerance of ambiguity, 85, 87–88
 Tone, 4–6, 15–17, 19, 28, 138, 152, 155, 303, 358, 359, 363, 410, 426–430, 433, 434, 449
 languages, 6, 16
 types, 433–434
TraciTalk, 255
 Transfer, 69, 74–77, 81, 88, 148, 186, 194, 386, 387, 437
 Translanguaging, 42, 124, 213, 214, 296
 Translation apps, 263
TriplePlayPlus, 247
 Tuning, 65, 193
 Turkish, 14, 191, 192, 325
Twitter, 237, 264, 274
 Two-part contrasts, 152, 362
Type and Talk, 370

U

Ultimate attainment, 70, 75, 77, 78, 80, 82, 83, 96, 332
 Undifferentiated gestures, 377, 378
 United Kingdom (UK), v, 19, 22, 48n5, 87, 128, 131, 348, 349, 352, 356, 357, 364, 371, 384, 385, 439, 440
 United States (U.S.), v, 12, 16, 19, 22, 48n5, 77, 78, 122, 123, 128, 156, 177, 178, 211, 257, 258, 312, 314, 320, 348, 349, 352, 356, 357, 364, 371, 372, 384, 389n2, 428

Universal Grammar (UG), 58, 83
 Unmarked, 32, 65, 76, 137
 Upspeak, 19, 23, 93, 440
 U.S. Army's Military Language Trainer (MILT), 257
 U.S. Medical Licensing Examination, 320
 Utterance fluency, 299, 304

V

Validation, 272, 290, 292, 293, 304, 313, 321, 328, 330, 333, 450, 451
 research, 323, 443–444
 Variationist sociolinguistics, 27
 Variegated babbling, 63
Versant, 179, 272, 273, 303, 321, 322, 324, 332
Ville, 254, 255
 Visible speech, 122, 309
VisiPitch IV, 240, 241
 Vocabulary, 60, 79, 180, 183–186, 191, 205, 210, 218, 253, 257, 260, 298, 322, 323, 350–352, 373, 383–385, 405–410
 Vocal styling, 358–360, 364, 366
 Voice
 biometrics, 367
 quality, vi, 5, 30–32, 65, 206, 207, 243, 244, 350, 357, 359, 360, 366, 370, 375, 389n3, 426
 recognition, 236, 245–247, 251, 252, 259, 371, 373
 VOICE (Vienna-Oxford International Corpus of English), 210
 Voice-activated GPS, 371

Voice-activated personal assistant
 apps, 264, 266, 371
Voice Changer, 244, 370
VoiceAssistant, 263
VoiceForge, 370
VoiceU, 4, 369
Voiceprint (spectrogram), 367
Voki, 370
Volume, 28, 97, 348, 351, 363, 366,
 428
Vowel, vi, 3–6, 10, 12, 14, 16, 25,
 28–30, 38, 40, 62, 63, 70, 79,
 129, 132, 137–141, 147, 148,
 151, 154, 155, 160, 161, 180,
 182, 184, 187, 201–203, 207,
 238, 240, 241, 251, 268, 272,
 297, 298, 353, 375, 383, 384,
 415, 417, 435, 438–440, 447,
 451n2, 451n3

W
Waveform, 236, 237, 242, 243, 246
Wavesurfer, 236
Web Tutorial, 239
West Africa, 150
WhatsApp, 237
Willingness to communicate, 87, 96,
 403, 412
WinPitchLTL, 240, 243
Words of the day, 264
Working class, 13, 26, 364, 365
Working memory, 60, 86, 89, 98,
 194, 385, 386
World of Warcraft (WoW), 257
Written L2, 185–186

Y
Ya-Ya Language Box, 249