Second Language Learning and Teaching

Veronica G. Sardegna
Anna Jarosz *Editors*

Theoretical and Practical Developments in English Speech Assessment, Research, and Training

Studies in Honour of Ewa Waniek-Klimczak

Springer

Second Language Learning and Teaching

Series Editor

Mirosław Pawlak, Faculty of Pedagogy and Fine Arts, Adam Mickiewicz University, Kalisz, Poland

The series brings together volumes dealing with different aspects of learning and teaching second and foreign languages. The titles included are both monographs and edited collections focusing on a variety of topics ranging from the processes underlying second language acquisition, through various aspects of language learning in instructed and non-instructed settings, to different facets of the teaching process, including syllabus choice, materials design, classroom practices and evaluation. The publications reflect state-of-the-art developments in those areas, they adopt a wide range of theoretical perspectives and follow diverse research paradigms. The intended audience are all those who are interested in naturalistic and classroom second language acquisition, including researchers, methodologists, curriculum and materials designers, teachers and undergraduate and graduate students undertaking empirical investigations of how second languages are learnt and taught.

More information about this series at https://link.springer.com/bookseries/10129

Veronica G. Sardegna · Anna Jarosz
Editors

Theoretical and Practical Developments in English Speech Assessment, Research, and Training

Studies in Honour of Ewa Waniek-Klimczak

Springer

Editors
Veronica G. Sardegna
Department of Instruction
and Leadership in Education
Duquesne University
Pittsburgh, PA, USA

Anna Jarosz
Faculty of Philology
Institute of English Studies
University of Lodz
Łódź, Poland

ISSN 2193-7648 ISSN 2193-7656 (electronic)
Second Language Learning and Teaching
ISBN 978-3-030-98217-1 ISBN 978-3-030-98218-8 (eBook)
https://doi.org/10.1007/978-3-030-98218-8

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This edited collection was written in honour of Professor Emeritus Ewa Waniek-Klimczak. She was Director of the Institute of English Studies and Head of the Department of English Language and Applied Linguistics at the University of Lodz in Łódź, Poland, until she retired in 2019. For many years, she has also been Chief Editor of *Research in Language*, an international journal publishing articles in linguistics, language acquisition, and human communication. During her academic career, Prof. Waniek-Klimczak published prolifically, co-authored many articles and book chapters, and co-edited many collections on the topics of phonetics, phonology, sociophonetics, L2 phonetics, and English pronunciation teaching. In 2007, she founded the annual International Conference on Native and Non-Native Accents of English 'Accents,' which has been growing steadily in popularity across the world since its inception.

'Accents' started as a modest conference and it soon turned into a truly international event gathering scholars from all parts of the world in Łódź every year in December. It has become a valuable forum for scientists and researchers interested in such areas as pronunciation, phonetics, discourse, prosody, accents in natural environments (World Englishes, immigrant contexts, sojourners, etc.) and educational contexts (ESL, EFL, ELF, ESOL, EAP, etc.), speech perception and speech production, and pronunciation instruction and instructed learning. Over fourteen years ('Accents 2022' is the fifteenth edition of the conference), this well-attended event has hosted the most distinguished plenary speakers. The Christmas atmosphere usually adds to the appeal and charm of the conference; however, it is undoubtedly Prof. Waniek-Klimczak's charismatic personality that has managed to create a friendly, positive, and welcoming ambience.

When the call for chapter proposals for this edited collection was sent out all over the world to scholars who know Prof. Waniek-Klimczak both professionally and privately, the response from potential contributors to this volume was extremely enthusiastic, which only indicates how well-known, respected, and distinguished she is among contemporary phoneticians and phonologists. Apart from Prof. Waniek-Klimczak's professional achievements, many referred to her kind, congenial, and helpful personality bringing back memories of past encounters, collaboration, and mutual projects. Even the busiest scholars responded that they would not miss on the opportunity to contribute to a volume in honour of Prof. Waniek-Klimczak. Thus, we received lots of proposals, which resulted in us being able to include chapters written by top scholars in the field and former 'Accents' plenary speakers, such as Pavel Trofimovich, Ron I. Thomson, Joan C. Mora, Steven H. Weinberger, Jan Volín, Katarzyna Dziubalska-Kołaczyk, Magdalena Wrembel, and Veronica G. Sardegna. The other contributions were written by exemplary researchers who also frequently present at 'Accents' conferences. All these well-known researchers came together to honour Prof. Waniek-Klimczak and to make the volume special.

This edited collection reflects the research areas Prof. Waniek-Klimczak has always been interested in. It covers both theoretical and practical approaches to English L2 speech assessment, research, and training. On behalf of all the authors, we would like to say that this volume is our tribute to our dearest friend and colleague, in recognition of her professional accomplishments as well as all the inspirational talks and most positive encouragement she is always ready to offer.

Pittsburgh, USA Veronica G. Sardegna
Łódź, Poland Anna Jarosz

Contents

Editors and Contributors

About the Editors

Veronica G. Sardegna, Ph.D. is Adjunct Faculty at Duquesne University in Pittsburgh, USA. She has taught ESL and teacher education courses at the University of Illinois at Urbana-Champaign, the University of Texas at Austin, the University of Pittsburgh, and Duquesne University, USA, for two decades. She conducts research on English pronunciation teaching, intercultural learning, and instructional technology. She has received the 2021 D. Scott Enright Interest Section Service Award for her outstanding service to TESOL.

Anna Jarosz is Assistant Professor in the Department of English Language and Applied Linguistics at the University of Lodz in Poland. Her professional interests include pronunciation teaching and learning with a focus on individual learner differences, motivation, and strategy use. She is the author of *English Pronunciation in L2 Instruction. The Case of Secondary School Learners* (Springer, 2019) and, since 2019, she organizes the International Conference on Native and Non-Native Accents ('Accents').

Contributors

Ali S. Alelaiwi Department of English, George Mason University, Fairfax, VA, USA

Tanja Kocjančič Antolík Institute of Phonetics, Charles University, Prague, Czech Republic

Gemma Archer University of Strathclyde, Glasgow, Scotland

Lina Bikelienė Vilnius University, Vilnius, Lithuania

Zuzanna Cal Adam Mickiewicz University, Poznań, Poland

Katarzyna Dziubalska-Kołaczyk Adam Mickiewicz University, Poznań, Poland

Ondřej Fischer Institute of Phonetics, Charles University, Prague, Czech Republic

Talia Isaacs UCL Centre for Applied Linguistics, London, UK

Anna Jarosz Faculty of Philology, Institute of English Studies, University of Lodz, Łódź, Poland

Pedro Luis Luchini Universidad Nacional de Mar del Plata, Mar del Plata, Provincia de Buenos Aires, Argentina

Takehiko Makino Faculty of Economics, Chuo University, Tokyo, Japan

Joan C. Mora Faculty of Philology and Communication, University of Barcelona, Barcelona, Spain

Ingrid Mora-Plaza University of Barcelona, Barcelona, Spain

Mary Grantham O'Brien University of Calgary, Calgary, Canada

Mireia Ortega University of Barcelona, Barcelona, Spain

Cosme Daniel Paz Universidad Nacional de Mar del Plata, Mar del Plata, Provincia de Buenos Aires, Argentina

Veronica G. Sardegna Department of Instruction and Leadership in Education, Duquesne University, Pittsburgh, PA, USA

Radek Skarnitzl Institute of Phonetics, Charles University, Prague, Czech Republic

Pavel Šturm Institute of Phonetics, Charles University, Prague, Czech Republic

Jolanta Sypiańska University of Szczecin, Szczecin, Poland

Kym Taylor Reid Concordia University, Montreal, Canada

Ron I. Thomson Brock University, St. Catharines, Canada

Pavel Trofimovich Concordia University, Montreal, Canada

Jan Volín Institute of Phonetics, Charles University, Prague, Czech Republic

Steven H. Weinberger Department of English, George Mason University, Fairfax, VA, USA

Ewelina Wojtkowiak Adam Mickiewicz University, Poznań, Poland

Magdalena Wrembel Adam Mickiewicz University, Poznań, Poland

Introduction

Introduction

Veronica G. Sardegna and Anna Jarosz

1 Why a Focus on English Speech Accents

Derwing and Munro (2015) define accent as “a particular pattern of pronunciation that is perceived to distinguish members of different speech communities” (p. 5). In dictionaries, it is frequently defined as a pronunciation style or manner identifying the country of origin, the geographical region, and the social background of its user. Interestingly, it is not possible to speak without an accent, which is inseparably linked with the concept of national, regional, ethnic, or social identity. Native accent variants can be treated as speech varieties that differ from each other in certain pronunciation features and characteristics. A foreign accent, frequently also referred to as accentedness, is a gradable phenomenon, which denotes the degree to which non-native speech differs from the particular native variety (Derwing & Munro, 2005).

English (L2) speech issues have gained in relevance with the spread of global international communication and the growing number of non-native English speakers who use English to communicate with other non-native English users—that is, as a lingua franca. As it has been attested in numerous studies, English L2 speakers have different objectives and needs with regard to their English pronunciation, including being understood by other users and/or by English natives, approximating a particular native accent, or hiding their own accent in order to come across as native speakers (Jarosz, 2019; Waniek-Klimczak, 1997). In L2 learning/teaching, nativeness seems

V. G. Sardegna (✉)
Department of Instruction and Leadership in Education, Duquesne University,
Pittsburgh, PA, USA
e-mail: sardegnav@duq.edu

A. Jarosz
Institute of English Studies, University of Lodz, Łódź, Poland
e-mail: anna.jarosz@uni.lodz.pl

V. G. Sardegna and A. Jarosz (eds.), *Theoretical and Practical Developments in English Speech Assessment, Research, and Training*, Second Language Learning and Teaching,
https://doi.org/10.1007/978-3-030-98218-8_1

to have become the secondary objective now that research points to intelligibility as a more feasible, attainable, and desired aim (Levis, 2018). Moreover, since intelligibility and accentedness are semi-independent constructs, considerably accented speech may still be understood by the interlocutor (Derwing & Munro, 1997). Thus, the degree of accentedness does not affect intelligibility or the delivery and reception of the intended message (Derwing & Munro, 1997).

However, the extent of foreign (L1) accentedness may impact the overall impression of listeners and bias their attitudes towards the speaker. It is worth bearing in mind that speech intelligibility and comprehensibility are not exclusively related to the speaker. The role of the listener or of the conversation topic (i.e., the degree of the listener's familiarity with it) cannot be underestimated (Rubin, 1992; Zielinski, 2008). The listener's attitude towards a particular accent (influenced by a particular L1 or representing a different region or background) affects the speech assessment and may evoke positive or negative prejudice in overall judgements (see Chaps. 2 and 3).

Therefore, English accents and accented speech, and the perception, production, and assessment of English speech varieties and pronunciation features are valuable areas to investigate further within the field devoted to L2 phonology acquisition. It is also important to examine the pedagogical effectiveness of various approaches to phonetic training and their effects on perceived L2 accent, comprehensibility, and intelligibility. Furthermore, there is also a growing demand for linguistically-grounded theories and models that would explain L2 and L3 acquisition, especially in the developing multilingual contexts. This volume was developed with these critical research needs in mind and with the goal of meeting the expectations of contemporary researchers and the requirements posed by modern linguistics.

2 Purpose, Aims, and Focus of the Book

The purpose of this edited collection is to present and discuss theoretical, practical, and research developments in English pronunciation in order to establish evidence-based directions and recommendations for best practices in English speech assessment, research, and training. The volume also provides a global perspective on English speech accents and their acquisition as well as pedagogical, assessment, and research implications for future research, including research directions for L2 and L3 acquisition.

Two main aims guided the selection of chapters: to disseminate knowledge about theoretical frameworks in relation to English speech assessment and training, and to share the results of the most recent investigations into L2 speech perception, production, assessment, and training. To achieve these aims, the volume features leading pronunciation experts and scholars who share valuable insights and cutting-edge research supported by contemporary methodologies and approaches.

The focus of this publication is on theoretical and practical frameworks for English speech assessment, perception, production, phonetic training, and acquisition. It highlights the importance of the role of the listener in speech assessment, of individual differences in moderating accent training, and of the links between accentedness and comprehensibility. It also addresses the needs of the growing multilingual contexts, in which English constitutes only one of the few acquired foreign languages. Thus, the approaches, methodologies, models, and theories presented in this volume aim to answer the most recent and most urgent questions posed in the field of English speech assessment, research, and training.

3 Intended Readers

This collection appeals to a large and mixed group of linguists, applied linguists, researchers, teachers, teacher educators, and students interested in English pronunciation learning, assessment, training, and research. Linguists, applied linguists, and researchers will find the book stimulating and thought-provoking because it not only confirms and disseminates the existing and available knowledge in the field of phonetics and phonology, but also sets new trends and directions for future research. The research and pedagogical methodologies and theories reported in the chapters will also undoubtedly motivate and encourage fruitful and intriguing academic discussions and exchanges among researchers in training and scholars.

Teachers and teacher educators may also appreciate the contributions of this volume as it offers them valuable insights into the mechanisms and factors that affect L2 speech acquisition, production, perception, and assessment. The chapters also share empirically-based theoretical frameworks and perspectives that can effectively inform and guide their teaching practices.

Undergraduate and postgraduate students could use this collection as a handbook for their courses in linguistics, second language acquisition, phonetics, and phonology. The book covers a carefully and methodically selected range of topics, research paradigms, and empirical and cutting-edge findings, which they may find useful, motivating, and relevant for their applied linguistics studies.

All in all, this volume is positioned to confirm, question, and suggest both theoretical and practical trends in English speech assessment, research, and training. Its international scope in terms of contributing authors and educational and research settings helps address the needs of a wide spectrum of readers, and makes it an enlightening and inspiring read for students and scholars at different levels of pedagogical and research expertise.

4 Structure of the Book

The collection consists of six main parts. Part I introduces the aims, the focus, and the structure of the book. It also comments on its intended readers.

Part II reviews, provides empirical evidence, and offers critical analyses guiding different aspects of English speech assessment, including the malleability of listener judgments, linguistic variables affecting listeners' reactions to accented speech, and the effects of prestige familiarity on students' perceptions and attitudes towards English speech. Kym Taylor Reid, Pavel Trofimovich and Mary Grantham O'Brien review the impact of negative and positive social bias on naïve listeners' and language teachers' L2 speech assessments. Then they suggest different ways of mitigating social bias effects on listeners' judgments and offer several implications for teaching, research, and assessment of L2 speech. Ron Thomson and Talia Isaacs focus on listeners' judgments of different L2 English speakers' fluency, comprehensibility, relaxedness, friendliness, and intelligence. They conclude with useful guidelines for L2 instruction and listener sensitivity training. Pedro Luchini and Cosme Paz analyze the usefulness of measurements of nuclear stress placement and comprehensibility in assessing L2 pronunciation. Gemma Archer investigates whether prestige accent varieties are still perceived prestigious in a blind test by international students. The final chapter by Takehiko Makino offers a critical analysis of several diagnostic passages commonly used for data collection and provides recommendations for read-aloud assessments.

Part III examines L2 listeners' perceptions of (a) accented/comprehensible non-native English speech, (b) two different strategies for syllable structure simplification, (c) voicing of English plosives in the coda position, and (d) stress of polysyllabic words. Joan Mora examines the relationship between accentedness and comprehensibility in L2 English speech as a function of non-native listeners' L1 and their L2 proficiency level. He also provides implications for L2 speech perception and assessment. Ali Alelaiwi and Steven Weinberger explore L2 perceptions of syllable structure simplifications consisting in consonant deletion and vowel epenthesis. Lina Bikelienė approaches the role of voicing of English plosive codas by Lithuanian learners of English from two perspectives: recognition and perception. Finally, Veronica Sardegna and Anna Jarosz report on a case study investigating the role of autonomous learning supported by *YouGlish* in predicting and perceiving English word stress in polysyllabic words.

Part IV reports empirical findings and research perspectives on the production of English vowels, coda obstruents, and rhoticity. Jan Volín, Tanja Kocjančič Antolík, Radek Skarnitzl and Pavel Šturm explore the factors that contribute to vowel accentedness in L2 learners. Steven Weinberger presents empirical evidence to support the claim that that /h/, aspiration, and vowel epenthesis are equivalent entities, which linguistically and theoretically equal the process of epenthesis in Mandarin Chinese production of coda obstruents. Ondřej Fischer and Pavel Šturm conclude this part with an investigation of the consistency of the treatment of rhoticity by Czech learners of English.

Part V shares current practices in phonetic training and their effect on learners and listeners. Ingrid Mora-Plaza, Mireia Ortega, and Joan Mora investigate the interplay between individual differences in auditory selective attention and attention switching skills, and the effectiveness of high-variability phonetic training under different stimuli and conditions in enhancing the perception and production of an

L2 vowel contrast. Finally, Ewelina Wojtkowiak shares the results of a longitudinal acoustic study exploring the effects of phonetic training on the acquisition of English stops by Polish learners.

Part VI presents theoretical perspectives on the acquisition of phonology in multilinguals. Katarzyna Dziubalska-Kołaczyk and Magdalena Wrembel propose the Natural Growth Theory of Acquisition to account for L2 and L3 acquisition processes. Jolanta Sypiańska and Zuzanna Cal explore perceptual drift in L1 phonetic categories caused by cross-linguistic influences from the L2 and L3.

References

Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition, 19*(1), 1–16. https://doi.org/10.1017/S0272263197001010S

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.* John Benjamins. http://dx.doi.org/10.1075/lllt.42

Jarosz, A. (2019). *English pronunciation in L2 instruction: The case of secondary school learners.* Springer. https://doi.org/10.1007/978-3-030-13892-9

Levis, J. M. (2018). *Intelligibility, oral communication, and the teaching of pronunciation.* Cambridge University Press. https://doi.org/10.1017/9781108241564

Rubin, D. (1992). Nonlanguage factors affecting undergraduates' judgements of nonnative English-speaking teaching assistants. *Research in Higher Education, 33*(4), 511–531. https://www.jstor.org/stable/40196047

Waniek-Klimczak, E. (1997). Context for teaching English phonetics and phonology at Polish universities and colleges: A survey. In E. Waniek-Klimczak & J. P. Melia (Eds.), *Accents and speech in teaching English phonetics and phonology: EFL perspective* (pp. 5–17). Peter Lang.

Zielinski, B. (2008). The listener: No longer the silent partner in reduced intelligibility. *System, 36*(1), 69–84. https://doi.org/10.1016/j.system.2007.11.004

Veronica G. Sardegna, Ph.D. is Adjunct Faculty at Duquesne University in Pittsburgh, USA. She has taught ESL and teacher education courses at the University of Illinois at Urbana-Champaign, the University of Texas at Austin, the University of Pittsburgh, and Duquesne University, USA, for two decades. She conducts research on English pronunciation teaching, intercultural learning, and instructional technology. She has received the 2021 D. Scott Enright Interest Section Service Award for her outstanding service to TESOL.

Anna Jarosz is Assistant Professor in the Department of English Language and Applied Linguistics at the University of Lodz in Poland. Her professional interests include pronunciation teaching and learning with a focus on individual learner differences, motivation, and strategy use. She is author of *English Pronunciation in L2 Instruction. The Case of Secondary School Learners* (Springer, 2019) and, since 2019, she organizes the International Conference on Native and Non-Native Accents (Accents).

Speech Assessment

The Malleability of Listener Judgments of Second Language Speech

Kym Taylor Reid, Pavel Trofimovich, and Mary Grantham O'Brien

Abstract In research and practice contexts, assessment of second language (L2) speakers often falls to listeners (e.g., naïve listeners, trained assessors, language teachers) who, as members of their respective linguistic communities, might be influenced by various social biases, both positive and negative. However, it is presently unclear whether listeners are immune to external social biases and how the potential impact of these biases on listener assessments of L2 speech can be mitigated. The goal of this chapter is therefore to review our recent empirical work examining the malleability of listeners' evaluations of L2 speech for several speech dimensions, including accentedness, comprehensibility, and fluency. We first review existing research regarding the impact of negative and positive social bias on naïve listeners' and language teachers' evaluations of L2 speech. We then discuss the roles of various interventions, such as task practice and rater training, as ways of mitigating social bias effects on listeners' speech assessments. We conclude by discussing possible implications of our findings for the teaching, research, and assessment of L2 speech.

Keywords Speech rating · Assessment · L2 English · Rating stability · Social bias · Task practice

K. Taylor Reid (✉) · P. Trofimovich
Concordia University, Montreal, Canada
e-mail: kym.taylor@concordia.ca

P. Trofimovich
e-mail: pavel.trofimovich@concordia.ca

M. G. O'Brien
University of Calgary, Calgary, Canada
e-mail: mgobrien@ucalgary.ca

V. G. Sardegna and A. Jarosz (eds.), *Theoretical and Practical Developments in English Speech Assessment, Research, and Training*, Second Language Learning and Teaching,
https://doi.org/10.1007/978-3-030-98218-8_2

1 Introduction

Social contact provides opportunities for people to form impressions about others. A person's speech, in particular, offers unique insight, providing listeners with various clues about the speaker's amiability, mood, place of origin, or education level (Yzerbyt et al., 1994). However, listeners' perceptions can also conflict with their expectations that originate from specific prior experiences or internalized beliefs. When such disparity occurs, listeners can either accept the new information or default to their initial beliefs, leading to an overgeneralization known as stereotyping or bias (Stroebe & Insko, 2013). This overgeneralization, based on unfounded beliefs and misinformation, can then be used as justification for marginalizing others. In our recent work, we have focused on socially constructed biases in the domain of second language (L2) speech to explore this societal problem. Our specific goal has been to determine the extent to which social biases occur in the context of listeners evaluating L2 speech, so that these biases can be reduced through various interventions, including perspective taking and task practice.

2 Literature Review

2.1 Origins of Social Bias

Humans naturally seek to identify themselves as members of various in-groups (i.e., groups to which they belong), often as a means of improving their chances of being perceived positively by others (Mullin & Hogg, 1999). This includes using speech (defined broadly as segmental and suprasegmental characteristics of speech contributing to listener perceptions of a speaker's accent) to categorize the speaker as being a member (or an outsider) of specific groups (Bourhis et al., 2012; Giles & Watson, 2013). In this way, a person's accent tells the story of who they are, where they have been (Matsuda, 1991) and, most importantly, where they belong. An example of this accent-based categorization can be found in Labov's (1972) classic study in which speakers showcased a unique identity as in-group members using their pronunciation (i.e., variation in vowel quality). More recent work has confirmed that people indeed use language to underscore between-group similarities and differences (and their associated stereotypes) and make judgments about others (Bourhis et al., 2012; Dragojevic et al., 2016; Ryan, 1983; Wigboldus et al., 2005).

Sometimes, speech-based distinctions are linked to positive listener attitudes. Speakers in Dalton-Puffer et al.'s (1997) study, for instance, who spoke British English with a Received Pronunciation accent, were rated by L2 learners as more courteous, educated, and organized than those with other native accents. Similarly, Heaton and Nygaard (2011) found that listeners from across the United States rated English speakers with an American southern accent to be more sociable than those with a standard American accent. When presented in isolation, these stereotypical attitudes do not appear to be particularly harmful but, in reality, the upgrading of one group often leads to the downgrading of another. In Heaton and Nygaard's study, for instance, though speakers with a southern accent were found to be more sociable than those with a standard American accent, the roles were reversed for ratings of intelligence. So, depending on which factors of the speakers are being assessed, someone is almost always in the out-group.

Negative attitudes that follow from such judgments can be particularly damaging when expressed by majority groups against minority status speakers, especially considering that humans exhibit a natural predisposition to "dislike the unlike" (Kagedan, 2020, p. 5). The detrimental effect of speech-based biases is perhaps best illustrated in workplace studies such as Ryan et al. (1977), where native English listeners perceived heavily-accented Spanish speakers to be of lower occupational status and to be less friendly than their less-accented counterparts. In fact, job applicants from minority groups are about 50% less likely to be invited to a job interview (Zschirnt & Ruedin, 2016), and immigrants (most of whom speak with a discernible foreign accent) tend to be underemployed (Krahn et al., 2000), with foreign accents considered more suitable for low than for high prestige jobs (Brennan & Brennan, 1981).

Attitudes toward specific cultural groups, societal norms regarding minority speakers, and the role of language within a particular society—including how it is used in education, politics, and the media—can all influence listener perception and stigmatization of L2 speakers (Gluszek & Dovidio, 2010). For instance, in an investigation of consumer responses to radio advertising, Lalwani et al. (2005) found that residents of Singapore believed standard English speakers to be more credible spokespersons than their fellow speakers of Singlish (Singapore English), even when pitching Singaporean products. To further complicate the issue of perception, at least some of these attitudes and behaviors can be attributed to listener expectations of speech before it is even heard (Lindemann & Subtirelu, 2013). In Babel and Russell's (2015) study investigating the effects of face priming on ratings of speech samples produced by native speakers of Canadian English, speech samples primed with photos of self-identified Chinese Canadians were rated as less intelligible and more accented than those primed with photos of self-identified White Canadians.

In another example in which expectations have been shown to fuel speech attitudes, Lindemann (2002) paired native English speakers with native Korean speakers for an interactive task in English. Some native speakers who were found to have negative attitudes toward Koreans prior to the interaction neglected to acknowledge communication from their partners and, in some cases, even withheld vital information. These same speakers subsequently reported their interactions to be less successful than native speakers who held positive attitudes toward Koreans. In fact, the root of most communication problems encountered during the task involved the "negative attitude" participants, which highlights the significant role of attitudes in successful communication, regardless of the language ability or proficiency of interlocutors.

When linguistic differences are used to judge speakers based on imagined or preconceived ideas, biases can become even more problematic. This is known as reverse linguistic stereotyping (Kang & Rubin, 2009), or the process by which general attributes of a speaking community negatively influence how a speaker is perceived, often based on completely imagined characteristics. In one example of this stereotyping, Rubin (1992) showed that native-speaking English listeners perceived an audio lecture paired with an image of a Chinese-looking female to be heavily accented, so much so that it interfered with the understanding of the lecture content. When the same audio was presented alongside an image of a Caucasian female, however, the content was understood significantly better and the speaker was rated as less accented, even though the audio was recorded, in both instances, by the same native English speaker from Ohio. Such preconceptions have also been found to affect the assessment of student work. For example, teachers of grade 3 and 4 schoolchildren provided higher evaluations of students' writing when it was paired with the speech of native English speakers than when students' writing was presented along with the speech of Spanish-accented speakers, regardless of which students actually produced the written work (Ford, 1984).

2.2 Manipulating Social Bias

Considering that listener-based evaluations of L2 speech are common in both research (e.g., Derwing & Munro, 2015) and assessment (e.g., Harding, 2012; Isaacs, 2013) contexts, it is important to thoroughly explore not only the effects of social biases on listener attitudes, but also to understand the conditions under which biases are amplified, as well as how biases can be mitigated toward better rating stability. For example, stereotypical yet extraneous information provided to participants can affect their performance. In a study conducted in an Italian-German bilingual community in Italy, Paladino et al. (2009) found that when Italian participants were simply reminded of the widely held perception that Italians of the region were known to have poor ability in German, they underperformed in their L2 German oral and written tests. Along the same lines, American listeners in Niedzielski's (1999) study perceived vowel sounds differently, even though they were produced by the same speaker, simply based on information that was provided prior to the listening task about the speaker being a resident of either the United States or Canada.

It is also possible to minimize existing biases. For instance, it is well documented that negative attitudes exist among native-speaking university students toward accented international teaching assistants (Halleck, 2008) and instructors (Hertel & Sunderman, 2009). Here, explicit training—in the form of increased exposure—can be useful in reducing bias. For instance, Staples et al. (2014) involved native-speaking undergraduate students in informal, cooperative contact activities with L2 speakers for eight weeks. Students who engaged in contact with L2 speakers subsequently rated L2 instructors more favorably on measures of accentedness (how closely the speaker approximates the target language variety), comprehensibility (how easy the speaker is to understand), and overall teaching ability than the group that had not taken part in additional cooperative tasks. Beyond explicit instruction, other mitigation tactics have also been successful in influencing socially constructed biases toward L2 speech. For instance, Hansen et al. (2014) invited German-speaking raters to walk in the shoes of an L2 speaker by conversing with a confederate researcher in the participants' own L2 (English) prior to rating the L2 German speech of Turkish speakers. Those who engaged in this form of perspective taking assigned higher ratings to the L2 speakers than raters who did not activate their L2 prior to the session.

3 Exploring Bias

Given that listener-based evaluations of L2 speech are influenced by extraneous factors, such as social biases and stereotypical views, these evaluations may not be as stable as originally thought. Nevertheless, when it comes to the assessment of L2 speech, human ratings are essential (Derwing & Munro, 2015). Teachers regularly evaluate L2 speakers in low-stakes assessments, such as oral presentations and tests, and trained assessors evaluate L2 speakers in higher-stakes contexts, such as standardized examinations. Naïve (untrained) listeners are also called upon often to rate speech for various dimensions, including accentedness and comprehensibility (Derwing & Munro, 2009). It is further common for untrained listeners to provide evaluations of L2 speakers that extend beyond speech itself. For example, individuals with no training in speech assessment have been asked to judge L2 speakers' socioeconomic status (Deprez-Sims & Morris, 2010), educational achievement (Campbell-Kibler, 2007), and competence (Baquiran & Nicoladis, 2020). Such evaluations often have implications for future work and study opportunities, wages, and the quality of healthcare that a person receives (Halim et al., 2017; Timming, 2017). If speech ratings are susceptible to social influences, then it is crucial to seek a better understanding of ways in which such rating-irrelevant variance (i.e., extraneous, uncontrolled variables that can influence assessment) can be minimized.

3.1 Social Bias and Naïve Listeners' Evaluations of L2 Speech

The initial objective of our work was to determine the effect of deliberate positive and negative social bias manipulation on naïve (untrained) listeners' ratings of L2 speech (Taylor Reid et al., 2019). This study, which explored the effects of social bias on listeners' assessments of native French speakers of L2 English from Quebec, Canada, was fueled by the idea that social influences, however slight, prior to the rating session might sway ratings assigned by naïve listeners relative to listeners who had not experienced any social influence prior to rating, especially in a social environment with a history of tension between English- and French-speaking communities.

Because the impact of social bias on listeners' judgments of L2 speech is likely determined by their specific experiences (e.g., Kang & Rubin, 2009; Wigboldus et al., 2005), the listeners recruited for this study represented a broad age range. We expected to find that older and younger listeners might differ in their ratings according to the impact of Quebec's language policy on their respective generations. The 1977 French Language Charter (Bill 101), which designated French as the sole official language of Quebec and restricted the use of English in public domains (including education) as a way of strengthening the ethnolinguistic vitality of francophones in Quebec (Corbeil, 2007), was expected to have heavily influenced the attitudes of older (40+) listeners. These listeners would have been children or young adults when the status of English changed from majority to minority, which would make them particularly sensitive to issues affecting English-speaking Quebecers. In contrast, younger listeners would have been raised and schooled at a time when the official status of French had been less contested, making them less sensitive to English-centered social influences. We therefore predicted that any social bias in listener evaluations would be qualified by listeners' age, leading to a more pronounced bias among older rather than younger listeners.

Sixty listeners were randomly assigned to one of three groups that engaged in a similar rating task, except for a brief personal story that a researcher shared with the listeners in two of the three groups at the outset of the session. In the negative manipulation group, 20 listeners (ages 19–66) heard negative comments by the researcher about a recent encounter with an L2 English speaker. In the positive manipulation group, 20 listeners (ages 18–72) heard a comparable opinion of the same length and emotional content reflecting the researcher's positive experience with an L2 English speaker. The 20 listeners (ages 20–65) in the baseline group rated the speech samples without any such manipulation. Regardless of group assignment, all listeners heard the same 40 brief L2 English narratives recorded by native French speakers from Quebec, assessing each speaker for accentedness, comprehensibility, and flow (overall pacing and speed of utterance delivery), as well as for specific pronunciation issues, such as segmental errors (accuracy in articulation of consonants and vowels) and intonation (natural rise and fall in pitch).

When the listeners were exposed to a positive bias manipulation, they generally behaved similarly in their ratings, irrespective of their age. Younger listeners upgraded the speakers for four of the five targeted measures (accentedness, comprehensibility, intonation, flow), while older listeners enhanced the speakers' evaluations for two measures (comprehensibility, intonation), compared to baseline listeners' assessments. However, the rating behaviors of the younger and older listeners diverged under a negative bias manipulation. Negatively oriented younger listeners pushed back against the researcher's negative comments, providing more favorable ratings for all five measures. However, this was not the case for the negatively oriented older listeners, who downgraded the same speakers relative to baseline listeners' evaluations. To put it another way, the researcher's biased comments about an L2 English speaker upset rating stability for the older listeners, who went along with the positivity (upgrading the speakers) and the negativity (downgrading the speakers), but the same comments functioned as an unexpected positive stimulus for the younger listeners under both the negative and positive bias manipulation conditions.

3.2 Social Bias and Teachers' Evaluations of L2 Speech

Armed with a clearer understanding of the influence of social bias on naïve listeners' evaluations of L2 speech, we then sought to investigate how that influence would affect expert raters in a new linguistic environment, this time with language teachers evaluating their students' performance (Taylor Reid et al., 2020). We specifically examined whether teachers of L2 German—as evaluators of their students' speaking performance—might also be sensitive to a social bias manipulation, just like the naïve listeners in our earlier work. We additionally investigated whether sensitivity to social bias might differ for teachers who are themselves either native speakers or non-native speakers of the language they are teaching.

With respect to native versus non-native teacher differences, it was possible that both native and non-native teachers of German would provide similar ratings for L2 speakers of German, regardless of the biasing orientation (e.g., Crowther et al., 2016; Derwing & Munro, 2013), because they are all part of the same professional group. However, negative comments made about L2 speakers might particularly resonate with the non-native teachers, invoking feelings of empathy that could carry over to more generous ratings (e.g., Hansen et al., 2014). Alternatively, the non-native teachers might show more negativity in their evaluations when exposed to social bias, given that L2 listeners sometimes provide harsher evaluations than native listeners (e.g., Kang, 2012; Rose, 2017; Rossiter, 2009).

In this study, conducted in the context of teaching and learning German as a foreign language in the English-speaking province of Alberta, Canada, we asked teachers of German to evaluate the speech of 24 intermediate to advanced L2 German speakers for the same five measures (i.e., accentedness, comprehensibility, flow, segmental errors, and intonation) as in our earlier study. Because teachers of L2 German in Alberta represent a relatively small group, which made it difficult to carry out a large-scale study, we recruited two comparable groups of teachers: (a) an experimental group made up of seven non-native and seven native teachers of German, and (b) a control group made up of seven non-native and seven native teachers of German. As in our earlier work, the two groups of teachers engaged in the same rating task, but only the experimental group received negative bias manipulation—that is, the researcher provided a negative opinion about the L2 German skills of a hypothetical learner of German prior to asking the teachers to rate the L2 German speech samples.

The native and non-native teachers provided comparable ratings of the intonation and fluency of L2 German speech, demonstrating similar rating behaviors for these speech dimensions. However, in response to negative bias, the native and non-native teachers diverged in their evaluations of the remaining three dimensions (accentedness, comprehensibility, and segmental errors). The native teachers downgraded the performance of L2 speakers, which corresponded to medium-strength statistical effect. In contrast, the non-native teachers provided more favorable evaluations for the same speakers. Put differently, the native teachers who heard negative comments about L2 students' German appeared to go along with the negativity, downgrading the speakers, whereas the non-native teachers seemed to show (enhanced) empathy with fellow L2 speakers, upgrading their ratings.

3.3 Summary and Outlook

Taken together, these two investigations shed new light on the stability of L2 speech ratings among both naïve and expert raters. In Taylor Reid et al. (2019), we found strong, consistent effects of positive and negative bias manipulations on all five targeted speech measures, such that the ratings provided by listeners under social bias diverged significantly from the ratings provided by baseline listeners. These findings add to the growing body of research in applied linguistics (e.g., Winke et al., 2013) and social psychology (e.g., Paladino et al., 2009) targeting various sources of bias in measures of L2 learning and use, and invite further investigations into social, attitudinal, and emotional underpinnings of listener assessments of L2 speech.

Our findings also cast doubt on the relative stability of human ratings of L2 speech, pointing to the importance of social context, defined both narrowly (as an immediate rating situation) and broadly (as a sociopolitical environment). Although the local context was tightly controlled, in that it took place in a laboratory, the real-world settings in which assessors find themselves before they pass judgment on L2 speakers are less rigidly controlled, which makes speech ratings particularly susceptible to various social influences. In fact, people are often unaware of the experiences that activate their preconceived ideas or stereotypes (Molden, 2014), so any negativity overtly or covertly attributed to L2 speakers can result in behaviors with important real-life consequences that extend, for example, to employment, wages, and healthcare (e.g., Halim et al., 2017; Hansen & Dovidio, 2016; Timming, 2017).

These concerns certainly extend beyond contexts involving naïve listeners. In Taylor Reid et al. (2020), we showed that language teachers are influenced by social bias, which differed in nontrivial ways in its influence on native versus non-native teachers. Language teachers evaluate L2 learners on a regular basis. Sometimes they carry out high-stakes evaluations that may determine, for example, whether a learner may study or work in a given target language setting. In such instances, it is essential that teachers be aware of both their own biases (as native speakers of the target language or as fellow non-native speakers) and the ways in which their assessments may be affected by comments provided by others. Importantly, too, employing multiple raters to carry out assessments in high-stakes settings may safeguard against the biases of individual raters. A prudent take-home message arising from this research is that not only are ratings more malleable than previously thought, but neither naïve listeners nor language teachers, as members of their respective sociolinguistic groups—whether they are laypersons or experienced language teachers, whether they are young or old, or whether they are native or non-native speakers themselves—are immune to social biases.

4 Mitigating Social Bias

Having established (at least some) effects of social bias on L2 speech ratings, we then explored ways to minimize rating-irrelevant social influences on evaluations of L2 speech. One mitigation strategy involves awareness raising through perspective taking (Boland & Tenkasi, 1995). Perspective taking refers to various activities whose goal is to guide people to consider various facets of another individual with whom they might surprisingly share commonalities. For instance, in a study of English-speaking university students' evaluations of L2 speakers, before eliciting speech judgments, Weyant (2007) asked some students to write about a day in the life of an L2 speaker, while other students were given no such instructions. The students who took the L2 speaker' perspective assigned her higher ratings of ability and accomplishment, compared to those who were not asked to write from the speaker's perspective. In another example, Zhang (2017) immersed pre-service music teachers in a 20-minute music class with all instructions and content delivered in Mandarin as a way of approximating the experience of learning the course content in another language. The teachers subsequently reported feeling anxious, confused, and frustrated, revealing an emotional response to perspective taking that may have allowed them to develop a new understanding of L2 speakers and their challenges.

4.1 Task Practice as Mitigation Strategy

Given that various forms of perspective taking appear to be successful at reducing listener bias (Weyant, 2007) or in creating greater awareness of the individuals being evaluated (Zhang, 2017), we reasoned that engaging listeners in task practice—essentially by asking them to perform the same speaking task as the speakers to be assessed—could stabilize listener behaviors by reducing social bias effects on their ratings. We explored this hypothesis in Taylor Reid et al. (2021).

As a starting point, we considered our finding from Taylor Reid et al. (2019): young English-French bilingual listeners (all dominant in English) were susceptible to negative and positive social bias in evaluation of L2 English speech. As discussed previously, the younger listeners exposed to both negative and positive bias manipulations upgraded L2 speakers significantly in their ratings, compared to the evaluations by those who were not exposed to a biasing social commentary, demonstrating rating-irrelevant variance in their speech assessments. With this finding in mind, in our most recent study (Taylor Reid et al., 2021), we targeted similar English-French bilinguals to examine whether engaging them in task practice in their more versus less dominant language could reduce rating-irrelevant variance in speech ratings arising through social bias. Because all social bias effects among younger listeners in our previous study led to inflated (more positive) ratings, we anticipated that task practice in listeners' more dominant language (English) would be more effective at minimizing social bias effects (i.e., bringing listeners' ratings more in line with the baseline listeners' ratings), compared to task practice in their less dominant language (French). In other words, those who are called upon to use their more dominant language prior to engaging in speech ratings might have higher expectations of L2 speakers, resulting in a reduction of any leniency that might be brought about by a biasing commentary. In contrast, the use of a less dominant L2 might be associated with rating leniency (e.g., Weyant, 2007), which may not be as effective at minimizing rating-irrelevant variance that has already been amplified through social bias.

Similar to the listeners in our earlier work, the 70 young bilingual listeners in this study were exposed to positively or negatively worded biasing comments. Most critically, however, before they provided their assessments of L2 speakers' accentedness and comprehensibility, 20 participants completed the same narrative task as the L2 speakers in their stronger language (English); another 20 participants performed the same task in their weaker language (French). The remaining 30 participants in the baseline groups did not engage in task practice and completed ratings of the 40 speech samples with negative, positive, or no social bias imposed. Only English task practice appeared to significantly reduce rating-irrelevant effects of social bias on listener assessments, and only under negative bias manipulation. In other words, those listeners who practiced the task in English were less likely to upgrade the speech of L2 speakers under the negative bias manipulation.

We reasoned that engaging the listener in the same speaking task completed by L2 speakers may have encouraged perspective taking in similar ways to writing about the life of a L2 speaker (Weyant, 2007). Performing the task in English rather than French was also useful, as it provided the listeners with a model to use in their evaluation of the speakers' L2 English speech, which is consistent with the positive role of increased task familiarity in rater training (Davis, 2016). In essence, increased task familiarity, along with realistic performance expectations available to the listeners through English task practice, may have limited the impact of the (negative) biasing commentary. A preliminary take-home message here is that task practice might be a medium through which L2 accentedness and comprehensibility ratings can be stabilized, thereby countering the effects of social bias.

4.2 Next Steps

In light of these findings, future work investigating human ratings of L2 speech should consider the impact of additional task practice and perspective taking interventions on listener-assessed dimensions of L2 speech. These could include reading or listening to anecdotes about situations where L2 speakers experienced prejudice or enhanced empathy from their interlocutors on the basis of their speech. Similarly, listeners could be asked to comment on situations in their own lives in which they experienced prejudice or enhanced empathy on the basis of their language performance or factors unrelated to their linguistic competence. If such interventions are effective in terms of encouraging more positive assessments of L2 speech in a laboratory setting, they could be utilized in the training of individuals tasked with assessing L2 speakers on a regular basis. For example, L2 teachers and examiners could be encouraged to complete tasks similar to those they are assessing before they begin their assessments.

Along similar lines, one aspect of training provided to human resources personnel could involve roleplaying that teases apart linguistic issues associated with people's speech from other factors related to their professional competence. In even higher-stakes contexts where assessment of credibility is paramount—such as those involving legal interaction with L2 speakers (as part of traffic stops, border crossings, courtroom proceedings)—efforts could move beyond conventional tactics aimed at increased understanding (e.g., diversity training) to those that might expand the mindset of the participant by combining perspective taking with other successful interventions such as intercultural communication opportunities, as can be achieved, for example, through virtual reality (see Salmanowitz, 2016).

5 Broader Implications, Future Work, and Conclusions

Collectively, our recent findings have enabled us to gain a clearer picture of how some forms of social bias, such as positively or negatively worded comments, and some types of interventions, such as task practice, can affect rater behavior, but tough work remains if we are to apply this knowledge to real-world contexts. For instance, mitigation found to be effective in a laboratory setting might naturally be extended to pedagogical contexts, where greater consistency and fairness in evaluations of L2 speech might be ensured through rater training, task familiarity exercises, and awareness-raising tactics that call attention to subconsciously held biases. The same findings might be applied to other organized contexts, such as the courtroom or corporate environment, where one might lean on diversity training and other tactics that guide decision-makers toward effective disentanglement of an L2 speakers' character and credibility from their linguistic status.

But what of the less-organized day-to-day interactions: traveler to traveler, teammate to teammate, neighbor to neighbor? To effect change in such circles, a more thorough exploration of the role of contextual factors—such as linguistic environment and exposure—on social attitudes toward L2 speech is warranted. We might also seek a sharpened understanding of how to best explain negative attitudes from a theoretical perspective so that we can maximize our efforts toward neutralizing negative attitudes at their roots. This could involve further examination through the lenses of such frameworks as communication accommodation and intergroup contact (Berry, 1997; Bourhis et al., 2012; Dragojevic et al., 2016). For instance, even if people enter a conversation with the best of intentions to collaborate toward effective and equitable communication, being able to control deep-seated biases and resist the urge to fall back on the upkeep and affirmation of one's social identity—often to the detriment of the interlocutor—is another matter.

More importantly, long-term mitigation of biases is likely dependent on targeted interaction. When the speaker and listener are separated—as in laboratory rater studies, scoring of university placement tests, or even listening to voicemail messages from prospective employees—there is no opportunity for interaction to occur. Furthermore, increased exposure to ethnic, linguistic, and cultural diversity might alleviate at least some negative biases in areas where regular contact with L2 speakers is limited. Such a finding could provide an indication that an individual's social network might be a contributing factor. If so, an investigation of community engagement opportunities that bring culturally diverse groups together in an appealing social context might be an excellent form of real-world mitigation. Additional research and practice might then focus on how effective such efforts are at permanently altering attitudes toward more equitable treatment of L2 speakers across contexts.

References

Babel, M., & Russell, J. (2015). Expectations and speech intelligibility. *The Journal of the Acoustical Society of America, 137*(5), 2823–2833. https://doi.org/10.1121/1.4919317

Baquiran, C. L. C., & Nicoladis, E. (2020). A doctor's foreign accent affects perceptions of competence. *Health Communication, 35*(6), 726–730. https://doi.org/10.1080/10410236.2019.1584779

Berry, J. W. (1997). Immigration, acculturation, and adaptation. *Applied Psychology: An International Review, 46*(1), 5–34. https://doi.org/10.1111/j.1464-0597.1997.tb01087.x

Boland, R. J., & Tenkasi, R. V. (1995). Perspective making and perspective taking in communities of knowing. *Organization Science, 6*(4), 350–372. https://doi.org/10.1287/orsc.6.4.350

Bourhis, R. Y., Sioufi, R., & Sachdev, I. (2012). Ethnolinguistic interaction and multilingual communication. In H. Giles (Ed.), *The handbook of intergroup communication* (pp. 100–115). Routledge. https://doi.org/10.4324/9780203148624.ch8

Brennan, E. M., & Brennan, J. S. (1981). Accent scaling and language attitudes: Reactions to Mexican-American English speech. *Language and Speech, 24*(3), 207–221. https://doi.org/10.1177/002383098102400301

Campbell-Kibler, K. (2007). Accent, (ING), and the social logic of listener perceptions. *American Speech, 82*(1), 32–64. https://doi.org/10.125/00031283-2007-002

Corbeil, J. C. (2007). *L'embarras des langues: Origine, conception et évolution de la politique linguistique québécoise*. Québec Amérique.

Crowther, D., Trofimovich, P., & Isaacs, T. (2016). Linguistic dimensions of second language accent and comprehensibility. *Journal of Second Language Pronunciation, 2*(2), 160–182. https://doi.org/10.1075/jslp.2.2.02cro
Dalton-Puffer, C., Kaltenboeck, G., & Smit, U. (1997). Learner attitudes and L2 pronunciation in Austria. *World Englishes, 16*(1), 115–128. https://doi.org/10.1111/1467-971X.00052
Davis, L. (2016). The influence of training and experience on rater performance in scoring spoken language. *Language Testing, 33*(1), 117–135. https://doi.org/10.1177/0265532215582282
Deprez-Sims, A. S., & Morris, S. B. (2010). Accents in the workplace: Their effects during a job interview. *International Journal of Psychology, 45*(6), 417–426. https://doi.org/10.1080/00207594.2010.499950
Derwing, T. M., & Munro, M. J. (2009). Comprehensibility as a factor in listener interaction preferences: Implications for the workplace. *Canadian Modern Language Review, 66*(2), 181–202. https://doi.org/10.3138/cmlr.66.2.181
Derwing, T. M., & Munro, M. J. (2013). The development of L2 oral language skills in two L1 groups: A 7-year study. *Language Learning, 63*(2), 163–185. https://doi.org/10.1111/lang.12000
Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research.* John Benjamins. https://doi.org/10.1075/lllt.42
Dragojevic, M., Gasiorek, J., & Giles, H. (2016). Communication accommodation theory. In C. R. Berger & M. E. Roloff (Eds.), *The international encyclopedia of interpersonal communication* (pp. 1–20). Wiley. https://doi.org/10.1002/9781118540190.wbeic006
Ford, C. E. (1984). The influence of speech variety on teachers' evaluation of students with comparable academic ability. *TESOL Quarterly, 18*(1), 25–40. https://doi.org/10.2307/3586333
Giles, H., & Watson, B. (Eds.). (2013). *The social meanings of language, dialect, and accent: International perspectives on speech styles.* Peter Lang. https://www.peterlang.com/document/1109208
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
Halim, M. L., Moy, K. H., & Yoshikawa, H. (2017). Perceived ethnic and language-based discrimination and Latina immigrant women's health. *Journal of Health Psychology, 22*(1), 68–78. https://doi.org/10.1177/1359105315595121
Halleck, G. B. (2008). The ITA problem: A ready-to-use simulation. *Simulation & Gaming, 39*(1), 137–146. https://doi.org/10.1177/1046878107308060
Hansen, K., & Dovidio, J. F. (2016). Social dominance orientation, nonnative accents, and hiring recommendations. *Cultural Diversity and Ethnic Minority Psychology, 22*(4), 1–8. https://doi.org/10.1037/cdp0000101
Hansen, K., Rakic, T., & Steffens, M. C. (2014). When actions speak louder than words: Preventing discrimination of nonstandard speakers. *Journal of Language and Social Psychology, 33*(1), 68–77. https://doi.org/10.1177/0261927X13499761
Harding, L. (2012). Pronunciation assessment. In C. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1–6). Wiley-Blackwell. https://doi.org/10.1002/9781405198431.wbeal0966
Heaton, H., & Nygaard, L. C. (2011). Charm or harm: Effect of passage content on listener attitudes toward American English accents. *Journal of Language and Social Psychology, 30*(2), 202–211. https://doi.org/10.1177/0261927X10397288
Hertel, T. J., & Sunderman, G. (2009). Student attitudes toward native and non-native language instructors. *Foreign Language Annals, 42*(3), 468–482. https://doi.org/10.1111/j.1944-9720.2009.01031.x
Isaacs, T. (2013). Assessing pronunciation. In A. J. Kunnan (Ed.), *The companion to language assessment* (pp. 140–155). Wiley-Blackwell. https://doi.org/10.1002/9781118411360.wbcla012
Kagedan, A. L. (2020). *The politics of othering in the United States and Canada.* Palgrave Macmillan. https://doi.org/10.1007/978-3-030-52444-9

Kang, O. (2012). Impact of rater characteristics and prosodic features of speaker accentedness on ratings of international teaching assistants' oral performance. *Language Assessment Quarterly, 9*(3), 249–269. https://doi.org/10.1080/15434303.2011.642631

Kang, O., & Rubin, D. L. (2009). Reverse linguistic stereotyping: Measuring the effect of listener expectations on speech evaluation. *Journal of Language and Social Psychology, 28*(4), 441–456. https://doi.org/10.1177/0261927X09341950

Krahn, H., Derwing, T., Mulder, M., & Wilkinson, L. (2000). Educated and underemployed: Refugee integration into the Canadian labour market. *Journal of International Migration and Integration/Revue de l'integration et de la migration internationale, 1*(1), 59–84. https://doi.org/10.1007/s12134-000-1008-2

Labov, W. (1972). On the mechanism of language change. In J. J. Gumperz & D. Hymes (Eds.), *Directions in sociolinguistics* (pp. 312–338). Holt, Rinehart, and Winston.

Lalwani, A. K., Lwin, M., & Li, K. L. (2005). Consumer responses to English accent variations in advertising. *Journal of Global Marketing, 18*(3–4), 143–165. https://doi.org/10.1300/J042v18n03_07

Lindemann, S. (2002). Listening with an attitude: A model of native speaker comprehension of nonnative speakers in the United States. *Language in Society, 31*(3), 419–441. https://doi.org/10.1017/S0047404502020286

Lindemann, S., & Subtirelu, N. (2013). Reliably biased: The role of listener expectation in the perception of second language speech. *Language Learning, 63*(3), 567–594. https://doi.org/10.1111/lang.12014

Matsuda, M. J. (1991). Voices of America: Accent, antidiscrimination law, and a jurisprudence for the last reconstruction. *Yale Law Journal, 100*(5), 1329–1407. https://doi.org/10.2307/796694

Molden, D. C. (2014). Understanding priming effects in social psychology: What is "social priming" and how does it occur? *Social Cognition, 32,* 1–11. https://doi.org/10.1521/soco.2014.32.supp.1

Mullin, B. A., & Hogg, M. A. (1999). Motivations for group membership: The role of subjective importance and uncertainty reduction. *Basic and Applied Social Psychology, 21*(2), 91–102. https://doi.org/10.1207/S15324834BA210202

Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology, 18*(1), 62–85. https://doi.org/10.1177/0261927X99018001005

Paladino, M.-P., Poddesu, L., Rauzi, M., Vaes, J., Cadinu, M., & Forer, D. (2009). Second language competence in the Italian-speaking population of Alto Adige/Südtirol: Evidence for linguistic stereotype threat. *Journal of Language and Social Psychology, 28*(3), 222–243. https://doi.org/10.1177/0261927X09335333

Rose, R. L. (2017). Differences in second language speech fluency ratings: Native versus nonnative listeners. In *Proceedings of the international conference "Fluency & Disfluency Across Languages and Language Varieties"* (pp. 101–103). Catholic University of Louvain. http://hdl.handle.net/2078.1/195807

Rossiter, M. J. (2009). Perceptions of L2 fluency by native and non-native speakers of English. *Canadian Modern Language Review, 65*(3), 395–412. https://doi.org/10.3138/cmlr.65.3.395

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

Ryan, E. B. (1983). Social psychological mechanisms underlying native speaker evaluations of non-native speech. *Studies in Second Language Acquisition, 5*(2), 148–159. https://doi.org/10.1017/S0272263100004824

Ryan, E. B., Carranza, M. A., & Moffie, R. W. (1977). Reactions toward varying degrees of accentedness in the speech of Spanish-English bilinguals. *Language and Speech, 20*(3), 267–273. https://doi.org/10.1177/002383097702000308

Salmanowitz, N. (2016). Unconventional methods for a traditional setting: The use of virtual reality to reduce implicit racial bias in the courtroom. *The University of New Hampshire Law Review, 15*(1), 118–160. https://scholars.unh.edu/unh_lr/vol15/iss1/2/

Staples, S., Kang, O., & Wittner, E. (2014). Considering interlocutors in university discourse communities: Impacting U.S. undergraduates' perceptions of ITAs through a structured contact program. *English for Specific Purposes*, *35*, 54–65. https://doi.org/10.1016/j.esp.2014.02.002

Stroebe, W., & Insko, C. A. (2013). Stereotype, prejudice, and discrimination: Changing conceptions in theory and research. In D. Bar-Tal, C. F. Graumann, A. W. Kruglanski, & W. Stroebe (Eds.), *Stereotyping and prejudice: Changing conceptions* (pp. 3–36). Springer. https://doi.org/10.1007/978-1-4612-3582-8_1

Taylor Reid, K., O'Brien, M. G., Trofimovich, P., & Bajt, A. (2020). Testing the malleability of teachers' judgments of second language speech. *Journal of Second Language Pronunciation, 6*(2), 237–265. https://doi.org/10.1075/jslp.19015.tay

Taylor Reid, K., Trofimovich, P., & O'Brien, M. G. (2019). Social attitudes and speech ratings: Effects of positive and negative bias on multiage listeners' judgments of second language speech. *Studies in Second Language Acquisition, 41*(2), 419–442. https://doi.org/10.1017/S0272263118000244

Taylor Reid, K., Trofimovich, P., & O'Brien, M. G., & Tsunemoto, A. (2021). Using task practice to reduce social influences on listener evaluations of second language accent and comprehensibility. *International Journal of Listening*. Advance online publication. https://doi.org/10.1080/10904018.2021.1904933

Timming, A. R. (2017). The effect of foreign accent on employability: A study of the aural dimensions of aesthetic labour in customer-facing and non-customer-facing jobs. *Work, Employment and Society, 31*(3), 409–428. https://doi.org/10.1177/0950017016630260

Weyant, J. M. (2007). Perspective taking as a means of reducing negative stereotyping of individuals who speak English as a second language. *Journal of Applied Social Psychology, 37*(4), 703–716. https://doi.org/10.1111/j.1559-1816.2007.00181.x

Wigboldus, D. H., Spears, R., & Semin, G. R. (2005). When do we communicate stereotypes? Influence of the social context on the linguistic expectancy bias. *Group Processes & Intergroup Relations, 8*(3), 215–230. https://doi.org/10.1177/1368430205053939

Winke, P., Gass, S., & Myford, C. (2013). 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

Yzerbyt, V. Y., Schadron, G., Leyens, J. P., & Rocher, S. (1994). Social judgeability: The impact of metainformational cues on the use of stereotypes. *Journal of Personality and Social Psychology, 66*(1), 48–55. https://doi.org/10.1037/0022-3514.66.1.48

Zhang, Y. (2017). Walking a mile in their shoes: Developing pre-service music teachers' empathy for ELL students. *International Journal of Music Education, 35*(3), 425–434. https://doi.org/10.1177/0255761416647191

Zschirnt, E., & Ruedin, D. (2016). Ethnic discrimination in hiring decisions: A meta-analysis of correspondence tests 1990–2015. *Journal of Ethnic and Migration Studies, 42*(7), 1115–1134. https://doi.org/10.1080/1369183X.2015.1133279

Kym Taylor Reid is an Applied Linguistics Ph.D. candidate in the Department of Education at Concordia University and a part-time faculty member at McGill University in Montreal, Canada. Her research focuses on social biases against non-native speech, social network analysis, sociolinguistic aspects of second language acquisition, and the teaching of second language pronunciation.

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

Mary Grantham O'Brien is Professor of German in the School of Languages, Linguistics, Literatures and Cultures at the University of Calgary, Calgary, Canada. Her research focuses on the perception and production of second language speech with an emphasis on suprasegmentals.

Evaluations of Foreign Accented Speech: Subjective Bias or Speech Signal Characteristics?

Ron I. Thomson and Talia Isaacs

Abstract It is often reported that native speakers negatively evaluate personality characteristics of second language (L2) speakers on the basis of their accent. Researchers have frequently concluded that such judgements result from implicit bias on the part of listeners, for whom an L2 accent triggers stereotypes about the ethnic or racial origin of the speaker. Within this research paradigm, little attention has been paid to quantifiable features of the speech signal that may also contribute to negative evaluations of accented speakers, independent of their ethnic or racial origins. A more detailed understanding of listener reactions to foreign accented speech, and the underlying linguistic variables that influence those reactions, can reveal complementary contributions of speakers and listeners to personal interactions. In this chapter, we report results from an empirical study in which 24 native-speaking undergraduate students rated monologic speech samples, produced by 36 L2 English speakers (18 Mandarin, 18 Slavic), for perceived fluency, comprehensibility, friendliness, intelligence, and listeners' comfort interacting with each speaker. Relationships between listener ratings, and quantitative speech measures were examined across speakers, and as a function of speakers' language background and speaking task. Results indicate that quantifiable features of L2 speech and task type influence reactions to foreign-accented speech.

Keywords Accent discrimination · L2 speech · Foreign accent · Comprehensibility · Social dimensions of accent · Stereotyping

R. I. Thomson (✉)
Brock University, St. Catharines, Canada
e-mail: rthomson@brocku.ca

T. Isaacs
UCL Centre for Applied Linguistics, London, UK
e-mail: talia.isaacs@ucl.ac.uk

V. G. Sardegna and A. Jarosz (eds.), *Theoretical and Practical Developments in English Speech Assessment, Research, and Training*, Second Language Learning and Teaching,
https://doi.org/10.1007/978-3-030-98218-8_3

1 Introduction

Experimental evidence has shown that listeners can detect a foreign accent after hearing as little as 30 milliseconds of second language (L2) speech (Flege, 1984). Listeners can also recognize an L2 accent in speech played backwards (Munro et al., 2010). Munro (2021) describes this skill as an artifact of general speech processing mechanisms. Scovel (1988) argues that these perceptual mechanisms originally evolved from a need to thwart unintentional mating with community outsiders. Recent research suggests that speakers of a majority language are less likely to form domestic partnerships with L2-accented speakers than they are with members of their own language community, and only slightly more likely to form friendships (Kogan et al., 2021). However, this appears to be a cultural rather than evolutionary phenomenon, since there is far less hesitation to marry someone of a different ethnicity if that person shares the same accent in some communities. At the furthest extreme, some argue that aversion to a foreign accent is triggered by racism (Ennser-Kananen et al., 2021).

While listener reactions to L2 speech can in part be influenced by cultural stereotypes formed in response to particular L2 accents, we take a more nuanced position. These reactions are also affected by speech signal characteristics. Non-native features, either transferred from a speaker's first language (L1) or developmental in nature, may affect listener reactions when they do not match listeners' previous linguistic experience. Even if an L2-accented utterance is intelligible, it may still place increased processing demands on listeners, potentially leading to their frustration and annoyance (Tulaja, 2020). These reactions should be seen as independent of cultural, ethnic and racial triggers. In this chapter, we present a study which demonstrates that factors under the control of L2 speakers can contribute to listener judgments. While not wanting to minimize the potentially deleterious effect of listeners' implicit or overt biases, identifying speaker-dependent variables can inform L2 pronunciation instruction that empowers learners to take greater control over the speech they produce, in view of promoting more positive listener reactions.

2 Literature Review

2.1 *Foreign Accent and Listener Bias Research*

While listener reactions to L2 accents can be both positive and negative, research focussing on negative reactions is largely predicated on a belief that humans are fundamentally prejudiced against those outside of their own speech community. Within this attitudinal research paradigm, L2 accent is seen as a salient feature which allows listeners to automatically activate personal biases, which then inform implicit evaluations of a speaker's social identity (e.g., Dewaele & McCloskey, 2015; Gluszek & Dovidio, 2010; Lev-Ari & Keysar, 2010; Lippi-Green, 2012; Shah, 2019).

Much of the research in this area relies on listener ratings of foreign-accented speech. Lambert et al.'s (1960) matched-guise technique has proven to be a popular procedure for obtaining these ratings. In this approach, listeners are asked to evaluate speech samples produced by bilingual or multilingual speakers of a target language who have recorded two or more versions of the same spoken text with different L1 and/or L2 accents. Recordings of multiple speakers are then randomized and played for listeners, who provide scalar judgments concerning inferred personal attributes and/or physical characteristics of the speakers. These attributes typically relate to speakers' status and/or solidarity, for example their social status or how friendly they are (Dragojevic & Goatley-Soan, 2020). While the matched-guise technique controls for voice quality by including paired utterances produced by the same speaker, eliciting samples in this way is not always practical or possible. Verbal-guise techniques with single L2-accented speech samples from each speaker are also used (Dragojevic & Goatley-Soan, 2020; Garrett, 2010). In these studies, the impact of any individual speaker's voice quality is assumed to average out over a large enough sample of L2 speakers.

Other techniques have also been used to elicit evidence of implicit bias towards L2-accented speakers. For example, even when a recorded speech sample is produced with a native accent, listeners can be prompted into perceiving an L2 accent if the recording is attributed to a picture of a foreign-looking speaker, a phenomenon that Kang and Rubin (2009) termed Reverse Linguistic Stereotyping (RLS). RLS research has demonstrated that a belief that a speaker is foreign can also impact native (Kang & Rubin, 2009; Rubin, 1992) and non-native (Ghanem & Kang, 2021) listeners' comprehension of native speaker productions. To find evidence of bias, Lindemann (2005) avoided using aural stimuli altogether, and instead had listeners describe accents associated with countries she presented via a map. This only required that listeners refer to personal recollections of L2 accents in order to make judgments about the personal attributes of speakers with those accents.

Regardless of the precise technique used, research investigating attitudinal reactions to foreign-accented speech consistently arrive at similar conclusions. Speaking with a foreign accent results in listeners downgrading L2 speakers' perceived social status and/or the degree of solidarity with the host or target language community (Dragojevic & Goatley-Soan, 2020; Ryan, 1983). Further, the extent to which an L2 speaker is downgraded reportedly depends on their ethnic origin. Lippi-Green (2012) and Lindemann (2005) argue that accents linked to non-Caucasian speakers are more likely to evoke negative evaluations than L2 accents associated with Caucasian speakers. Gilchrist and Chevrot (2017) demonstrate that explicit ethnic attribution, in which listeners are made aware of the ethnic background of the speaker, impacts assessment of speakers' global L2 proficiency. Specifically, judges assigned lower proficiency scores to Arabic-accented English speakers than to Portuguese- and Chinese-accented English speakers on the basis of speech samples that contained exactly the same content.

While attitudinal research demonstrates that an L2 accent can trigger biases, which can then unduly affect listener reactions to the speakers, the strength of this association may be exaggerated. First, these studies tend to be highly controlled and therefore lack ecological validity. In the real world, listeners' attention may not be as explicitly oriented towards a speaker's ethnicity or social identity as laboratory findings suggest. In fact, many listeners are not even able to accurately identify L2 accented speakers' ethnicity in the lab (Dragojevic & Goatley-Soan, 2020; Gilchrist & Chevrot, 2017; Lindemann, 2003). Second, these studies rarely discuss individual differences across listeners, but instead focus on differences in group means. Not all listeners within a population sample respond in the same way (Dewaele & McCloskey, 2015; Kang & Yaw, 2021). Third, there is little overall focus on linguistic features stemming from the speech signal that may trigger positive or negative reactions quite apart from biases related to presumed group identity of L2-accented speakers. Thus, by convention, attitudinal research treats accent as a unidimensional, global phenomenon.

2.2 Impact of L2 Accents on Speech Processing

While attitudinal research has largely focussed on listener bias, examining ways in which particular features of L2 accents impinge upon speech processing by listeners can provide a complementary and richer account for negative reactions (e.g., Kang, 2012). Among researchers interested in L2 pronunciation learning and teaching, there is a widely-established literature evidencing multiple partially independent dimensions along which listeners respond to L2 speech (Derwing & Munro, 2015; Thomson, 2018). Munro and Derwing (1995a, 1995b) draw a distinction between foreign accent (in relation to a target norm), comprehensibility (listeners' perception of effort in processing speech), and intelligibility (how much listeners actually understand). Foreign accent on its own may evoke listener bias, but poor comprehensibility and/or intelligibility can also trigger negative reactions independent of any assumptions about a speaker's ethnicity.

Numerous L2 pronunciation studies have demonstrated that distinctive phonological features of L2 speech affect listener perceptions of accent, comprehensibility, and intelligibility in different ways (Derwing & Munro, 2015; Isaacs & Thomson, 2020; Kang et al., 2010, 2020; Levis, 2018; Munro & Derwing, 2006). While most L2 phonological features comprise negative transfer from the learners' L1s, there is considerable individual variation across speakers who share the same L1. L2 features can also be developmental in nature, reflecting interlanguage patterns which may impact speakers across a variety of L1 backgrounds. Segmental, prosodic, and temporal features of an L2 accent which are incongruent with a given listener's experience can cause processing difficulties. Derwing et al. (2009), for example, found that Mandarin L2 English speakers transferred L1 vowel length patterns to L2 English, negatively impacting their speech rate, which correlates with listeners' perception of fluency. Because the nature and extent of incongruence varies depending on each learner's L1, some L2 accents are more challenging for listeners to process than are others. While the attitudinal research described earlier reports that non-Caucasian L2 English accents are downgraded more than Caucasian L2 English accents, is this solely the result of group bias, or does a greater phonological distance between

English and specific non-Caucasian languages (e.g., Mandarin) contribute to this effect? The effect of L1-L2 phonological distance and strength of L2 accent has been considered in numerous pronunciation studies (Bongaerts et al., 2000; Bradlow et al., 2010; Cristia et al., 2012; Isaacs & Thomson, 2020), but not to our knowledge in attitudinal research.

In the temporal domain, perceived oral fluency of L2 speakers has also been shown to interact with listener perceptions of accentedness, comprehensibility, and intelligibility (Derwing & Rossiter, 2003; Derwing et al., 2004; Thomson, 2015). Attitudinal researchers have recognized the impact of perceived fluency on listener reactions to L2-accented speech (Alter & Oppenheimer, 2009; Dovidio & Gluszek, 2012). However, they seem to construe it as a shortcoming on the part of listeners, rather than something for which speakers bear some responsibility. In an effort to demonstrate a triggering effect of L2 identities on perceived fluency, Dragojevic and Goatley-Soan (2020) did not account for measurable differences in the L2 speakers' speech rate or proficiency. It is quite possible that those whose scores were downgraded by listeners were both less fluent and less proficient.

Other lines of research have similarly failed to consider the contribution of L2 speech processing difficulties on listener judgments. Pantos and Perkins (2013) used an implicit association test to demonstrate that response latencies to negative words were faster when associated with a foreign accent than with a native accent. However, is it the foreignness of the accent that is the issue, or might processing difficulty trigger negative emotions, which then drive listeners' association with negative words? Romero-Rivas et al. (2016) argue that difficulties listeners experience in anticipating upcoming words during sentence processing of L2 speech are caused by listeners activating negative affect in response to a speaker's accent. Could it not also be the case that properties of a particular accent are challenging, which would lead to identical results? Even evidence from neurolinguistics is used to support implicit biases. Foucart and Hartsuiker (2021) found differences in neurological activity during a sentence processing task when listeners were asked to judge the truth-value of true/false statements produced by native versus non-native speakers. They took these differences be an indicator of negative bias. Might these differences not simply be related to the greater effort that is sometimes required to process L2-accented speech? Others have found that familiarity with an L2 accent mitigates adverse reactions (Dewaele & McCloskey, 2015; Kang & Yaw, 2021). While they conclude that familiarity decreases bias, the fact that greater familiarity also leads to faster processing suggests that the appearance of bias might not be caused by the identity of the speaker but by psycholinguistic limitations on the part of the listener.

2.3 Task Type

The nature of speaking tasks used in most attitudinal research makes determining the underlying causes of negative reactions to L2 accents difficult. Typically, this research relies upon highly controlled tasks to elicit predictable speaker output. While control is important for making comparisons across speakers, controlled tasks may not reflect reactions in real world communication. For example, a highly decontextualized reading task (e.g., Dragojevic & Goatley-Soan, 2020) might interact with particular accents to induce a larger negative affect than is otherwise warranted. Heaton and Nygaard (2011) found that the specific content of a passage can affect listener attitudes. This negative effect might not be replicated in response to potentially more engaging spontaneous speech. Another concern is that once a controlled passage is known to listeners, they will be better able to attend to pronunciation features of the speech sample, potentially increasing the saliency of accents. Listening to repetitive speaking tasks may also contribute to rater fatigue, which might be amplified in the case of foreign-accented samples.

3 The Study

While we do not dispute the contribution of linguistic stereotyping and discrimination as factors in how listeners evaluate L2-accented speech, we do not believe that it provides a complete account. In this exploratory investigation, we set out to determine if other factors might account for much of the variance in negative evaluations of L2-accented speech.

3.1 Research Questions

In this study, we examine reactions to L2 English speech samples produced by Mandarin and Slavic-accented speakers, performing two speaking tasks. We consider the influence of quantifiable features of the speech signal (i.e., speech rate and pitch) as they relate to listeners' perceptions of comprehensibility and fluency, and their inferences about speakers' personality characteristics. Specifically, we asked:

1. Is there any evidence of bias in listener reactions to Mandarin vs. Slavic-accented English speakers?
2. Are listeners' judgments influenced by temporal and/or prosodic properties of L2 speakers' productions?
3. Do differences in L2 speaking performance across tasks contribute to different listener judgment patterns?

3.2 Method

3.2.1 L2 English Speakers

The L2 speech data were elicited from 36 adult newcomers to Canada who were enrolled in a government-funded ESL program. Half were Chinese (Mandarin) L1 speakers and half were Slavic L1 speakers (mostly Russians, but also two Ukrainians, two Serbians, a Pole, and a Serbo Croatian). Apart from L1 differences, the groups comprised similar demographics. All were identified as beginners according to the Canadian Language Benchmarks 3–4 (CLB levels 1–4 of the instrument; Pawlikowska-Smith, 2000). The Mandarin group included 14 females and 4 males (*M* age 40.2 years; range 29–49). The Slavic group included 12 females and 6 males (*M* age 38.8 years; range 27–47). The mean age of first English exposure was 14 years (range 12–36) for the Mandarin group and 16 years (range 1–33) for the Slavic group, with most close to the mean. The L2 speakers' self-reported use of English outside of the classroom was also comparable, with approximately 1/3 of their daily communication in English, and more time spent watching English TV/videos than anything in their L1. One notable difference between the groups was in their estimated daily interactions with NSs outside of the classroom. The Mandarin group averaged less than one hour per day, while the Slavic group averaged nearly two hours per day.

3.2.2 L2 Speaking Task

We used an eight-frame picture description task as well as a personal narrative task to elicit L2 speech. The picture sequence illustrates a humorous event in which a man and a woman mix up their identical suitcases after bumping into each other on a city street. This story has been widely used in previous research (Derwing et al., 2004; Isaacs & Thomson, 2013, 2020). In the personal narrative task, participants were asked to describe their experiences during the first two weeks after their arrival to Canada. Recordings of all speaking tasks were made in a quiet room using a digital recorder, paired with a high quality unidirectional Sennheiser microphone. While recording length varied across participants, the picture description task was usually completed in less than 2 minutes, while the personal narrative tasks typically lasted between 2 and 3 minutes. Following Isaacs and Thomson (2013, 2020), we only used the first 20 seconds of each speaking task, after removing any initial false starts or other initial dysfluencies. This resulted in 144 items (36 speakers x 4 tasks). We created three randomizations of these items for presentation to raters. In each, we interspersed three recordings of native speakers completing the first picture description task. The native speaker items were used to ensure that the raters were scoring the correct speech sample, since we anticipated that the native speakers would receive high scores.

3.2.3 NS English Listeners

Twenty-four native English speaker listeners (21 female, 3 male) provided ratings of the L2 speech samples. All were undergraduate social science students at a mid-sized English-medium Canadian university (*M* age 22.8, range 19–49). Most were monolingual, although seven self-reported being fluent in a second language. All had spent the majority of their lives in Ontario, six were from the Toronto metropolis, and 20 from smaller cities. All reported normal hearing. None of the listeners had any previous formal experience rating L2-accented speech.

3.2.4 Rating Task

Three rating sessions were conducted in a quiet room, each with a group of eight raters. They were presented with one of three randomizations of the L2 recordings via loudspeaker and were asked to rate each speaker for fluency and comprehensibility during a first session, and friendliness, intelligence and how comfortable they were interacting with the speaker in a second session. They recorded their assessments on printed paper using 9-point Likert-type scales. Scales and their endpoints were as follows:

Fluency: Very dysfluent—very fluent
Comprehensibility: Very hard to understand to very easy to understand
Friendliness: Not very friendly—very friendly
Intelligence: Not very intelligent—very intelligent
Interactional comfort: Not very comfortable—very comfortable

We provided raters with very brief instructions at the beginning of the rating sessions, explaining that comprehensibility refers to how easy it is to understand a speaker, while fluency refers to how smooth the speaker's oral delivery is based on their use of pauses, hesitations, fillers, etc. We gave no guidance on how to interpret friendliness and intelligence as we took these to be subjective constructs. For the final category, we simply asked, 'How comfortable would you feel interacting with this person?' After giving these instructions, we had the group of raters listen to two examples and discuss together how they might rate the samples on the relevant scales. We also told the listeners that all L2 speech was produced by speakers of Chinese or Slavic origin. Other than this brief introduction, raters were encouraged to indicate responses based on their subjective assessments of each speech sample. At the end of the second session, raters completed a short questionnaire with fixed and open-ended questions that asked them to elaborate on factors that influenced their evaluation of each speech construct on each task.

3.2.5 Discrete Measures of Speech

In addition to obtaining listener ratings, we extracted quantitative measures of the samples' temporal and acoustic characteristics. We used pruned syllables per second as a measure of speech rate. This was operationalized as the total speaking time divided by the number of fluent syllables produced (i.e., we did not count syllables comprising self-corrections, self-repetitions and nonlexical fillers such as 'um'). Among a wide variety of common speech rate measures, Derwing et al. (2004) found the pruned syllable measure to be the most strongly correlated with listener judgments of fluency. Total speaking time was measured using Sound Studio 3 and pruned syllables were calculated with reference to transcripts that had been created by a research assistant and verified by the first author.

We also calculated each speaker's minimum and maximum pitch (in Hz) and pitch range over the duration of each speaking task as a marker of affect (Ohala, 1983). Ohala found that higher pitch is associated with friendliness and politeness, while lower pitch is associated with confidence and dominance. Pitch measures were extracted using Praat (Boersma & Weenink, 2016). Automatic pitch tracks were used in a first pass, and manually corrected in some instances where the pitch tracker failed. After extracting pitch values for each speaker, male values were normalized to the female mean in order to combine data across all speakers.

3.3 Results

3.3.1 Interrater Reliability

We calculated interrater reliability for each rating scale using Cronbach's Alpha. Scores evidenced high overall consistency across raters as follows: Fluency (0.94), Comprehensibility (0.93), Friendliness (0.88), Intelligence (0.91), Interactional comfort (0.90).

To examine L1 and task differences we computed a series of five partially repeated-measures ANOVAs, one for each speech/personality construct. Speaking task (2 levels) served as a within-subject factor, while L1 was a between-subject factor. Results (see Table 1) indicate a significant difference in ratings when comparing performance on the picture description versus the personal narrative task. Across all scales, speaker performance on the personal narrative task was always rated more favorably than on the picture description task (see Fig. 1). For fluency, comprehensibility, intelligence and interactional comfort scales, Slavic-accented speakers were rated more positively than Mandarin-accented speakers, with small to medium effect sizes. For the friendliness scale, however, there was no significant difference between Slavic and Mandarin-accented speakers.

To examine L1 and task differences for speech rate (pruned syllables/sec) and pitch range measures, we computed two partially repeated measures ANOVAs, one for each quantitative measure. Speaking task (2 levels) served as a within-subject

Table 1 Results of partially repeated measures ANOVAs comparing mean ratings for each task by L1group

	Task			L1		
	$F(1,34)$	p	$\eta 2$	$F(1,34)$	p	$\eta 2$
Fluency	10.020	0.003	0.228	11.064	0.002	0.246
Comprehensibility	11.017	0.002	0.245	19.050	< 0.001	0.359
Friendliness	11.939	0.001	0.260	0.089	0.768	0.003
Intelligence	22.455	< 0.001	0.398	9.703	0.004	0.222
Interaction comfort	14.617	< 0.001	0.301	10.399	0.003	0.234

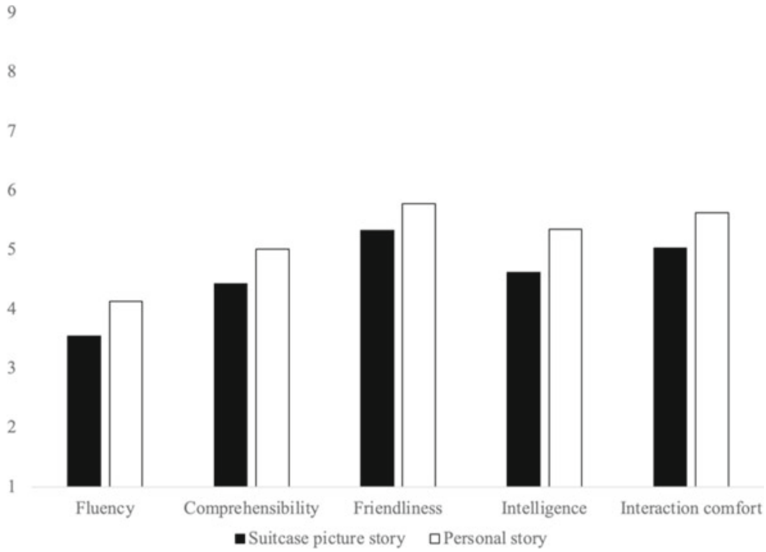

Fig. 1 Mean speech and personality scale ratings by task

Table 2 Results of partially repeated measures ANOVA comparing speech rate and pitch measures X L1group

	Task			L1		
	$F(1,34)$	p	$\eta 2$	$F(1,34)$	p	$\eta 2$
Speech rate (pruned syllables/sec)	10.047	0.003	0.228	0.682	0.415	0.020
Max pitch	4.869	0.034	0.125	3.367	0.075	0.090
Mean pitch range	5.515	0.025	0.140	4.357	0.044	0.114

factor, while L1 was a between-subject factor. Results (see Table 2) indicate a significant difference in speech rate, with a faster rate on the personal narrative task compared to the picture description task. No L1 effect for speech rate was detected, however. For maximum pitch, a significant effect was found for Task, with the picture description having higher maximum pitch ($M = 348$ Hz) than the personal narrative ($M = 321$ Hz). No significant difference was found for L1. Speakers used a significantly larger pitch range on the picture description task ($M = 271$ Hz) relative to the personal narrative task ($M = 243$ Hz). Further, Slavic speakers used a significantly larger pitch range ($M = 274$ Hz) than Mandarin speakers ($M = 239$ Hz).

3.3.2 Multiple Regression Analyses

Stepwise linear multiple regression analyses were conducted to examine to what extent temporal characteristics of the produced speech (pruned syllables/sec) and pitch measures (maximum, minimum, and range) predicted ratings on each task. On the suitcase picture description task (see Table 3), pruned syllables/sec were

Table 3 Multiple regression of variables contributing to listener reactions to the picture description task

	Predictors Stepwise (R^2)	Standardized coefficients (β)	t value	p value	Partial correlation
Fluency	Pruned syllables (0.677)	0.823	8.439	< 0.001	0.823
Comprehensibility	Pruned syllables (0.810)	0.900	12.052	< 0.001	0.900
Friendliness	Max pitch (0.147) Pruned syllables (0.317)	0.449 0.418	3.084 2.871	0.004 0.007	0.473 0.447
Intelligence	Pruned syllables (0.644) Max pitch (0.736)	0.851 0.307	9.392 3.392	< 0.001 0.002	0.853 0.508
Interaction comfort	Pruned syllables (0.520) Max pitch (0.670)	0.783 0.392	7.729 3.873	< 0.001 < 0.001	0.721 0.269

Table 4 Multiple regression of variables contributing to listener reactions to personal narrative task

	Predictors Stepwise (R^2)	Standardized coefficients (β)	*t* value	*p* value	Partial correlations
Fluency	Pruned syllables (0.436)	0.660	5.124	< 0.001	0.660
Comprehensibility	Pruned syllables (0.241)	0.491	3.287	0.002	0.491
Friendliness	Max pitch (0.187) Pruned syllables (0.349)	0.454 0.403	3.228 2.863	0.003 0.007	0.490 0.446
Intelligence	Pruned syllables (0.463)	0.680	5.415	< 0.001	0.680
Interaction comfort	Pruned syllables (0.337)	0.580	4.154	< 0.001	0.580

strong predictors of fluency and comprehensibility. A combination of pruned syllables/second and maximum pitch range strongly predicted intelligence and interactional comfort ratings, but only weakly predicted friendliness ratings. On the personal narrative task (see Table 4), pruned syllables were weaker predictors of fluency, comprehensibility, intelligence, and interactional comfort. Maximum pitch along with pruned syllables/sec combined to weakly predict friendliness ratings on the personal narrative task.

In addition to linear regression analyses, descriptive Pearson correlation coefficients also revealed significant relationships between pitch range and friendliness ratings on the picture description task (0.373) and personal narrative task (0.403). No significant correlations were found with minimum or maximum pitch, however.

Looking at simple correlation plotlines, we observed a few outliers that did not fit the overall patterns. Thus, we examined the transcripts of four samples comprising the greatest mismatch between attitudinal ratings and related quantitative measures. Following Heaton and Nygaard (2011) we found some evidence that content may have played a role in the evaluation of speakers' friendliness and intelligence. For example, Sp45 was the most monotone of all speakers, yet was rated as the sixth friendliest (out of the 36) on the personal narrative task. At the beginning of his story, this speaker said, "Canada uh knocked off my socks. Haha. When I arrived in Canada people was very friendly." Such content may have caused positive affect among raters. Sp38 was the 35th most monotone, yet was rated as the 11th friendliest on the personal narrative. In his story he states, "Our family immigrated to Canada… I met the some people. They are very nice and kindness." We found similar mismatches between intelligence ratings and quantitative measures. Sp21 was rated as the 9th most intelligent, despite being ranked 27th in terms of pruned syllables/sec. The speaker said, "when I came here, I didn't know nobody. I, I must rent apartment, and meet new people and find, must find work." The content seems to describe a confident and/or independent person. Conversely, Speaker 37 was rated as 27th in terms of intelligence, despite producing the 12th highest pruned syllables/second. She related, "at first I was very sad and nobody helped me and where the going. I no idea."

3.3.3 Questionnaire Data

Responses to the fixed-choice portion of our post-rating questionnaire also revealed how particular features of the oral texts may have contributed to listener reactions. Most raters (88%) indicated that even if individual words were clearly pronounced, a lack of coherence in the stories affected their ratings. Nearly two-thirds (63%) indicated being impacted by how fluently a speaker proceeded through the story, while almost all (92%) indicated positive affect when a speaker spent time developing details of a story (e.g., the "beautiful city" or "tall buildings"). The same number (92%) were negatively impacted by incorrect word choice for important words (e.g., "bit each other" instead of "bumped into each other").

Responses to an open-ended question asking for raters' top two influences on their ratings also revealed a diversity of influences. Some (38%) explicitly referenced fluency as a determinant in their ratings. One noted that "fluency was very important to delivery" while another stated that fluency might reflect it being "harder to tell the [picture story] than to tell their own personal experience." Many raters (42%) commented on how easy or difficult it was to understand the speech samples. One indicated that "speaking clearly and pronouncing words correctly made the story easier to understand" while another stated, "more cohesion means it's easier to communicate using Standard English." Another rater said, "if I couldn't understand them I had a more difficult time listening to their story." Only a few raters made comments related to influences on friendliness ratings (12.5%) and intelligence ratings (12.5%). One suggested that how much the speaker "enjoyed telling [the story]" influenced how friendly they sounded, while another pointed out that monotone speech "didn't seem as friendly." One rater stated that "the story sounded better if [speakers] knew what they were talking about," while another said "when they sound intelligent the story is easier to listen to."

4 Discussion

Traditional attitudinal researchers might interpret our listeners' scalar ratings as providing evidence that listeners activate implicit (or explicit) biases that associate Caucasian, Slavic-accented speakers with greater intelligence, interactional comfort, fluency, and comprehensibility relative to non-Caucasian, Mandarin-accented speakers (Lindemann, 2005; Lippi-Green, 2012). Differences in quantifiable features of the speech samples provide important insight, however. While we found no measurable L1-based differences in speech rate, Slavic speakers had a significantly wider pitch range than the Mandarin speakers. Interestingly, this itself is contrary to stereotypes that Slavic speakers are monotone (Crosby, 2013; Svetozarova, 1998) and to evidence that pitch range in Mandarin is much wider than that of English (Chen, 1974). Multiple regression analyses revealed that a combination of pruned syllables and speakers' maximum pitch combined to strongly predict intelligence and interactional comfort ratings, and that pruned syllables account for much of the variance in fluency and comprehensibility ratings. This suggests that much of the difference in attitudinal reactions favouring Slavic accented speakers in our study are attributable to differences in how each L1 group controls L2 English speech rate and pitch. While there were no overall differences in how friendly Slavic and Mandarin speakers were perceived to be, maximum pitch and pitch range influenced friendliness ratings across both groups. Finally, we found that ratings were higher on all scales for the personal narrative task than for the picture description task. Examination of the content of L2 oral productions and rater questionnaire data added further nuance, suggesting that the nature of the speaking task plays a crucial role in ratings and that some tasks allow speakers to create solidarity with raters by the things that they say during the task (e.g., that they are happy in Canada). This suggests that the content of an L2 utterance or their overall message could affect listeners' judgments of L2 speech, unduly resulting in higher or lower ratings when this is extraneous to the L2 speaking construct being measured (e.g., fluency, comprehensibility).

5 Implications

Like Derwing and Munro (2015), we see communication as a two-way street. In their primary research context involving immigrant language learning, they are rightly concerned that not enough emphasis is given to listeners' responsibility in accommodating L2 accents. In attitudinal research, however, the opposite seems to be the case. Little attention is paid to what speakers can do to make themselves more intelligible. In our exploratory study, we found evidence of listener bias in reactions to L2-accented speech, but also contributions from quantifiable features of the speech signal produced by learners. We only examined two features of speech, temporal fluency and pitch, however. There are many more segmental and prosodic features that might influence listener reactions (Derwing & Munro, 2015; Kang et al., 2010, 2020). To the extent to which L2 learners want to (e.g., McCrocklin & Link, 2016) and are able to change features of their pronunciation that induce negative reactions by listeners, they should be encouraged to do so, just as listeners should be encouraged to become more tolerant with L2 accented speakers. The good news is that L2 pronunciation instruction can be quite effective (Thomson & Derwing, 2015), and if aimed at improving speakers' intelligibility or comprehensibility (rather than global accent), it is often worth the effort. One means to help listeners is to encourage them to have more interaction with L2-accented speakers, since there is some evidence that familiarity makes processing more efficient (Derwing & Munro, 2014; Porretta et al., 2017). Familiarity can also lead to measurably less bias in listener responses to L2-accented speakers (Dewaele & McCloskey, 2015; Kang & Yaw, 2021).

6 Conclusion and Limitations

The results of this study suggest that there is a need for more research to tease apart the relative impacts of attitudinal bias versus quantifiable L2 speech characteristics that influence reactions to foreign-accented speakers. Listeners' reactions to positive versus negative framing of the host community's culture in an L2 speech sample is one area that would benefit from further examination in both controlled experiments, and in more in-depth qualitative data. One limitation in the current study is its lack of a native accented comparison group. Further, we did not consider listener familiarity with Mandarin and Slavic accents.

References

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review, 13*(3), 219–235. https://doi.org/10.1177/1088868309341564

Boersma, P., & Weenink, D. (2016). *Praat: Doing phonetics by computer* (Version 6.0.19) [Computer Software]. http://www.praat.org

Bongaerts, T., Mennen, S., & van der Slik, F. (2000). Authenticity of pronunciation in naturalistic second language acquisition: The case of very advanced late learners of Dutch as a second language. *Studia Linguistica, 54*(2), 298–308. https://doi.org/10.1111/1467-9582.00069

Bradlow, A., Clopper, C., Smiljanic, R., & Walter, M. A. (2010). A perceptual phonetic similarity space for languages: Evidence from five native language listener groups. *Speech Communication, 52*(11–12), 930–942. https://doi.org/10.1016/j.specom.2010.06.003

Chen, G. T. (1974). The pitch range of English and Chinese speakers. *Journal of Chinese Linguistics, 2*(2), 159–171.

Cristia, A., Seidl, A., Vaughn, C., Schmale, R., Bradlow, A., & Floccia. C. (2012). Linguistic processing of accented speech across the lifespan. *Frontiers in Psychology, 3,* 1–15. https://doi.org/10.3389/fpsyg.2012.00479

Crosby, C. F. (2013). *L1 influence on L2 intonation in Russian speakers of English* (Publication No. 1070). Master's thesis, Portland State University. Dissertation and Theses. https://doi.org/10.15760/etd.1070

Derwing, T. M., & Munro, M. J. (2014). Training native speakers to listen to L2 speech. In J. M. Levis & A. Moyer (Eds.). *Social dynamics in second language accent* (pp. 219–236). De Gruyter Mouton.

Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. John Benjamins.

Derwing, T. M., Munro, M. J., Thomson, R. I., & Rossiter, M. J. (2009). The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition, 31*(4), 533–557. https://doi.org/10.1017/S0272263109990015

Derwing, T. M., & Rossiter, M. J. (2003). The effects of pronunciation instruction on the accuracy, fluency, and complexity of L2 accented speech. *Applied Language Learning, 13*(1), 1–17.

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

Dewaele, J. M., & McCloskey, J. (2015). Attitudes towards foreign accents among adult multilingual language users. *Journal of Multilingual and Multicultural Development, 36*(3), 221–238. https://doi.org/10.1080/01434632.2014.909445

Dovidio, J. F., & Gluszek, A. (2012). Accents, nonverbal behavior, and intergroup bias. In H. Giles (Ed.), *The handbook of intergroup communication* (pp. 109–121). Routledge.

Dragojevic, M., & Goatley-Soan, S. (2020). Americans' attitudes toward foreign accents: Evaluative hierarchies and underlying processes. *Journal of Multilingual and Multicultural Development*. Advance online publication. https://doi.org/10.1080/01434632.2020.1735402

Ennser-Kananen, J., Halonen, M., & Saarinen, T. (2021). "Come join us and lose your accent!": Accent modification courses as hierarchization of international students. *Journal of International Students, 11*(2), 322–340. https://doi.org/10.32674/jis.v11i2.1640

Flege, J. E. (1984). The detection of French accent by American listeners. *Journal of the Acoustical Society of America, 76*(3), 692–707. https://doi.org/10.1121/1.391256

Foucart, A., & Hartsuiker, R. J. (2021). Are foreign-accented speakers that 'incredible'? The impact of the speaker's indexical properties on sentence processing. *Neuropsychologia*, 107902. https://doi.org/10.1016/j.neuropsychologia.2021.107902

Garrett, P. (2010). *Attitudes to language*. Cambridge University Press.

Ghanem, R., & Kang, O. (2021). ESL students' reverse linguistic stereotyping of English teachers. *ELT Journal, 75*(3), 330–340. https://doi.org/10.1093/elt/ccab011

Gilchrist, C., & Chevrot, J. P. (2017). Snap judgment: Influences of ethnicity on evaluations of foreign language speaking proficiency. *Corela: Cognition, Représentation, Langage, 15*(1). https://doi.org/10.4000/corela.4920

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

Heaton, H., & Nygaard, L. C. (2011). Charm or harm: Effect of passage content on listener attitudes toward American English accents. *Journal of Language and Social Psychology, 30*(2), 202–211. https://doi.org/10.1177/0261927X10397288

Isaacs, T., & Thomson, R. I. (2013). Rater experience, rating scale length, and judgments of L2 pronunciation: Revisiting research conventions. *Language Assessment Quarterly, 10*(2), 135–159. https://doi.org/10.1080/15434303.2013.769545

Isaacs, T., & Thomson, R. I. (2020). Reactions to second language speech: Influences of discrete speech characteristics, rater experience, and speaker first language background. *Journal of Second Language Pronunciation, 6*(3), 402–429. https://doi.org/10.1075/jslp.20018.isa

Kang, O. (2012). Impact of rater characteristics and prosodic features of speaker accentedness on ratings of international teaching assistants' oral performance. *Language Assessment Quarterly, 9*(3), 249–269. https://doi.org/10.1080/15434303.2011.642631

Kang, O., & Rubin, D. L. (2009). Reverse linguistic stereotyping: Measuring the effect of listener expectations on speech evaluation. *Journal of Language and Social Psychology, 28*(4), 441–456. https://doi.org/10.1177/0261927X09341950

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

Kang, O., Thomson, R. I., & Moran, M. (2020). Which features of accent affect understanding? Exploring the intelligibility threshold of diverse accent varieties. *Applied Linguistics, 41*(4), 453–480. https://doi.org/10.1093/applin/amy053

Kang, O., & Yaw, K. (2021). Social judgement of L2 accented speech stereotyping and its influential factors. *Journal of Multilingual and Multicultural Development*, 1–16. Advance online publication. https://doi.org/10.1080/01434632.2021.1931247

Kogan, I., Dollmann, J., & Weißmann, M. (2021). In the ear of the listener: The role of foreign accent in interethnic friendships and partnerships. *International Migration Review, 55*(3), 746–784. https://doi.org/10.1177/0197918320988835

Lambert, W. E., Hodgson, R., Gardner, R. C., & Fillenbaum, S. (1960). Evaluational reactions to spoken languages. *Journal of Abnormal and Social Psychology, 60*(1), 44–51. https://doi.org/10.1037/h0044430

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

Levis, J. M. (2018). *Intelligibility, oral communication, and the teaching of pronunciation*. Cambridge University Press.

Lindemann, S. (2003). Koreans, Chinese, or Indians? Attitudes and ideologies about non-native English speakers in the United States. *Journal of Sociolinguistics, 7*(3), 348–364. https://doi.org/10.1111/1467-9481.00228

Lindemann, S. (2005). Who speaks "Broken English"? US undergraduates' perceptions of non-native English. *International Journal of Applied Linguistics, 15*(2), 187–212. https://doi.org/10.1111/j.1473-4192.2005.00087.x

Lippi-Green, R. (2012). *English with an accent: Language, ideology and discrimination in the United States* (2nd ed.). Routledge.

McCrocklin, S., & Link, S. (2016). Accent, identity, and a fear of loss? ESL students' perspectives. *Canadian Modern Language Review, 72*(1), 122–148. https://doi.org/10.3138/cmlr.2582

Munro, M. J. (2021). *Applying phonetics: Speech science in everyday life*. Wiley-Blackwell.

Munro, M. J., & Derwing, T. M. (1995a). 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. (1995b). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech, 38*(3), 289–306. https://doi.org/10.1177/002383099503800305

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., & Burgess, C. S. (2010). Detection of nonnative speaker status from content-masked speech. *Speech Communication, 52*(7–8), 626–637. https://doi.org/10.1016/j.specom.2010.02.013

Ohala, J. J. (1983). Cross-language use of pitch: An ethological view. *Phonetica, 40*(1), 1–18. https://doi.org/10.1159/000261678

Pantos, A. J., & Perkins, A. W. (2013). Measuring implicit and explicit attitudes toward foreign accented speech. *Journal of Language and Social Psychology, 32*(1), 3–20. https://doi.org/10.1177/0261927X12463005

Pawlikowska-Smith, G. (2000). *Canadian Language Benchmarks 2000: Theoretical framework*. Centre for Canadian Language Benchmarks.

Porretta, V., Tremblay, A., & Bolger, P. (2017). Got experience? PMN amplitudes to foreign-accented speech modulated by listener experience. *Journal of Neurolinguistics, 44,* 54–67. https://doi.org/10.1016/j.jneuroling.2017.03.002

Romero-Rivas, C., Martin, C. D., & Costa, A. (2016). Foreign-accented speech modulates linguistic anticipatory processes. *Neuropsychologia, 85,* 245–255. https://doi.org/10.1016/j.neuropsychologia.2016.03.022

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

Ryan, E. B. (1983). Social psychological mechanisms underlying native speaker evaluations of non-native speech. *Studies in Second Language Acquisition, 5*(2), 148–159. https://doi.org/10.1017/S0272263100004824

Scovel, T. (1988). *A time to speak: A psycholinguistic inquiry into the critical period for human speech.* Newbury House.

Shah, A. P. (2019). Why are certain accents judged the way they are? Decoding qualitative patterns of accent bias. *Advances in Language and Literary Studies, 10*(3), 128–139. https://doi.org/10.7575/aiac.alls.v.10n.3p.128

Svetozarova, N. (1998). Intonation in Russian. In D. Hirst & A. di Christo (Eds.), *Intonation systems: A survey of twenty languages* (pp. 261–274). Cambridge University Press.

Thomson, R. I. (2015). Fluency. In M. Reed & J. Levis (Eds.), *The handbook of pronunciation* (pp. 209–226). Wiley. https://doi.org/10.1002/9781118346952.ch12

Thomson, R. I. (2018). Measurement of accentedness, intelligibility and comprehensibility. In O. Kang & A. Ginther (Eds.), *Assessment in second language pronunciation* (pp. 11–28). Routledge.

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

Tulaja, L. (2020). Exploring acceptability: L1 judgements of L2 Danish learners' errors. In O. Kang, S. Staples, K. Yaw, & K. Hirschi (Eds.), *Proceedings of the 11th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Northern Arizona University, September 2019 (pp. 197–206). Iowa State University.

Ron Thomson is Professor of Applied Linguistics/TESL at Brock University, Canada. His research focuses on the development of L2 pronunciation and fluency, listener responses to foreign-accented speech, and ethical practice in pronunciation instruction. Dr. Thomson is also known for his work in computer-mediated instruction, and how it may facilitate easier and more rapid development of L2 speech perception and production.

Talia Isaacs is Associate Professor of Applied Linguistics and TESOL at the UCL Institute of Education, University College London (UCL), UK, and Programme Leader for the MA TESOL In-Service. Her research broadly aims to reduce language barriers and improve the quality and fairness of oral communication assessment for English language learners.

Assessing L2 Pronunciation Using Measurements of Nuclear Stress Placement and Comprehensibility

Pedro Luis Luchini and Cosme Daniel Paz

Abstract Nuclear stress in English highlights the most important information in a sentence. Its correct use and location are thus fundamental for achieving meaningful communication. English learners who manifest intelligibility and/or comprehensibility problems due to nuclear stress misplacement can improve their pronunciation through explicit focused instruction. This classroom-based study aimed to evaluate the effectiveness of two pronunciation instruction treatments in an EFL context using measurements of nuclear stress placement and comprehensibility. Participants were 50 Spanish-L1 trainees divided evenly into Groups A and B. Both groups were exposed to a traditional, teacher-centered approach to pronunciation teaching (TCT), but Group B added a communicative, awareness-building component (CABC). Participants' free speech samples were assessed before and after instruction via pre- and post-test recordings. A slight tendency for improvement for nuclear stress and higher values for comprehensibility were observed between pre- and post-tests for Group B. A statistically significant simple linear regression was reported only for Group B in the relative response for nuclear stress and comprehensibility, thus demonstrating the benefits of CABC. The assessment protocols proved useful in determining the efficacy of one treatment over the other. The chapter concludes with a discussion of the implications of the findings for pronunciation assessment, research, and teaching.

Keywords L2 pronunciation assessment · Nuclear stress placement · Comprehensibility

P. L. Luchini (✉) · C. D. Paz
Universidad Nacional de Mar del Plata, Mar del Plata, Provincia de Buenos Aires, Argentina

V. G. Sardegna and A. Jarosz (eds.), *Theoretical and Practical Developments in English Speech Assessment, Research, and Training*, Second Language Learning and Teaching,
https://doi.org/10.1007/978-3-030-98218-8_4

1 Introduction

For years, second language (L2) pronunciation instruction has been marginalized in L2 language teaching and research. To date, however, with the advent of economic globalization, technological advances and the subsequent need to establish effective oral communication, pronunciation teaching has been revitalized and has thus reemerged in the applied linguistic research mainstream. As a result, there is a large number of high-ranking pronunciation-specific studies that recognize explicit instruction as crucial for the development of learners' L2 speech intelligibility and comprehensibility (e.g., Derwing et al., 2014; Lee et al., 2015; Saito, 2011; Trofimovich & Isaacs, 2012; Trofimovich et al., 2017). Despite this renewed interest in L2 pronunciation teaching, many in-service and pre-service teachers still report being confused about how to teach and/or assess this construct (Foote et al., 2011; Murphy, 2014; Pennington & Rogerson-Revell, 2019) unless their teacher education programs specifically trained them for pronunciation teaching and assessment (Sardegna, 2020). Also, while some studies have found that segmental aspects are crucial for effective communication, particularly those that carry a high functional load (Munro & Derwing, 2006), others provide robust evidence that prosody, particularly nuclear stress, is essential for understanding (Dauer, 2005; Hahn, 2004; Jenkins, 2000; Luchini, 2017), and should thus be explicitly taught. Therefore, an exploration of the relationship between measurements of nuclear stress placement and comprehensibility and instructional approaches may provide useful information regarding the efficacy of L2 pronunciation instruction.

This chapter reports on an experimental study that conducted such exploration with 50 trainees at a local university in Mar del Plata, Argentina. Participants' first language (L1) was Spanish and their L2 was English. They were split in two groups depending on the pronunciation instruction approach they received. Two specialists assessed their pre- and post-instruction speech samples with respect to nuclear stress placement and ten English native speakers judged the comprehensibility of the same speech samples. The results obtained provide interesting implications for L2 pronunciation assessment, research and teaching, which are discussed at the end of the chapter.

2 L2 Pronunciation Teaching and Assessment

The last 20 years have witnessed a paradigm shift in the goals of L2 pronunciation teaching as numerous renowned research studies in the field have given precedence to intelligibility and comprehensibility over those of nativeness or the eradication of a foreign accent (Derwing & Munro, 2009, 2015; Levis, 2005, 2018). Research findings have shown that pronunciation is a vital component of communicative competence and as such it should be given high priority in the L2 classroom (Morley, 1991). Provided that learners are intelligible and easy to understand, their pronunciation will not obstruct communication. Levis (2005) classified L2 pronunciation teaching into two main categories: teaching that follows the *Nativeness Principle* and teaching that follows the *Intelligibility Principle*. The first category comprises pronunciation instruction whose main goal is to push students to achieve a native-like pronunciation, while the second refers to speech that listeners can comfortably understand despite having some traces of local or regional accent coming from the speaker's L1.

Munro and Derwing (1995) and Derwing and Munro (1997) presented three different dimensions of L2 speech. They refer to *intelligibility* as the extent to which a listener understands L2 speech. They define *comprehensibility* in regard to the measure of how easy or difficult it is for the listener to understand L2 speech; that is, the cognitive effort required by the listener to understand. Lastly, they define *accentedness* as differences between speakers' and listeners' speech production of sounds and sound patterns. Accent is partly independent from intelligibility and comprehensibility (Trofimovich & Isaacs, 2012). Although accent is perceptively evident, it does not necessarily obstruct understanding (Derwing & Munro, 2009, 2015). Frequently, it is difficulties with intelligibility and/or comprehensibility that may cause problems for understanding. This suggests that the main goal for L2 pronunciation teaching should be to focus on those features of pronunciation that may cause problems for understanding, unless the learner's speech is already very clear.

Many studies examine the comparative efficacy of L2 pronunciation teaching for either segmental (sounds) or suprasegmental features (stress, rhythm and intonation). While various researchers show that sounds are easier to teach and learn (Levis, 2005; Saito, 2014), others claim that suprasegmental-based instruction has an influence on comprehensibility (Derwing et al., 1998; Gordon et al., 2013; Hahn, 2004; Isaacs & Trofimovich, 2012; Kang et al., 2010; McNerney & Mendelsohn, 1992; Munro, 1995; Saito & Saito, 2017). As for the different aspects that make up the suprasegmentals in English, it is known that the protagonist is nuclear stress. This salient prosodic feature plays a decisive role in producing textual cohesion and in sequencing a hierarchical organization of discourse. Nuclear stress points to new and contrastive information and data that are not available for the listener to retrieve from the context or prior knowledge (Bardovi-Harlig, 1986; Halliday, 2013; Pennington & Richards, 1986; Sperber & Wilson, 1986).

Many L2 learners have difficulty learning how to use nuclear stress in English. They often display two major problems: stressing almost all words in an utterance without signaling one major prominent stress, and/or misplacing nuclear stress (Field, 2005; Hahn, 2004). When nuclear stress is misplaced, sentence processing for the listener becomes more difficult, thus compromising comprehensibility (Birch & Clifton, 1995; Kang et al., 2010; Tajima et al., 1997; Terken & Hirschberg, 1994; Winters & O'Brien, 2013). Non-native speakers' intonation, for example, seems to be a crucial factor in native listeners' understanding, as tone choice and location can affect both perceived information structure and pragmatic cues in L2 discourse (Kang et al., 2010).

Numerous research studies show that prosodic features have a strong impact on L2 oral performance assessment (Anderson-Hsieh et al., 1992; Derwing et al., 1998; Isaacs & Trofimovich, 2012; Kang, 2012; Kang & Johnson, 2018; Kang et al., 2010; Saito, 2014). While most of these investigations explore different linguistic and phonological variables and their correlates with intelligibility and comprehensibility, few of them assess the effects of single parameters of English such as nuclear stress placement (Kang & Johnson, 2018) and their relationship with comprehensibility as predictors of L2 oral development. Therefore, classroom-based studies evaluating the properties of such prosodic features and their relationship with comprehensibility warrant further exploration.

In the past, the assessment of L2 pronunciation was marginalized in second language teaching and research, mainly because it was associated with discrete aspects of oral discourse (Lado, 1961). The advent of communicative competence (Hymes, 1972) brought about the years of neglect of pronunciation. It was not until the mid-90s that teachers and researchers began to focus their attention on the value and role of L2 pronunciation assessment for effective language use. At present, the emergence of pronunciation assessment can be partially ascribed to the shift in focus from perceptions of accentedness to the wide-ranging L2 speech dimensions of intelligibility and comprehensibility.

Recent indications among researchers and educational practitioners show that pronunciation assessment has attracted particular interest and gained special importance (Bøhn & Hansen, 2017; Chun, 2006, 2008; Fulcher, 2015; Isaacs, 2008, 2016; Isaacs & Trofimovich, 2016; Kang & Pickering, 2014; Kim, 2015; Thomson, 2018; Trofimovich et al., 2016; Xi, 2010, 2012), signaling that this process is now being considered part of the L2 speaking construct. Until recently, research on pronunciation assessment has relied heavily on listeners' subjective judgments and other L2 speech measurements external to the listener such as speech rate, pause length and location, lexical stress, among others. However, there have been few attempts to explore the impact of measurements of nuclear stress placement and comprehensibility working in tandem for the development of L2 speech.

3 The Study

This classroom-based study sought to assess L2 learners' pronunciation using measurements of nuclear stress placement and comprehensibility to evaluate the efficacy of one particular pronunciation pedagogical treatment over another. The research questions that guided the present study are:

1. To what extent does the addition of a communicative, awareness-rising component to a traditional teacher-centered approach to the teaching of L2 pronunciation contribute to enhance the students' perceived comprehensibility and nuclear stress placement?
2. Is there any degree of association between comprehensibility and nuclear stress placement improvements in each of the treatments applied?

4 Method

4.1 Context and Participants

The experimental context of this classroom-based study was *Discurso Oral II* (DOII), a 16-week pronunciation-specific course of eight weekly hours focusing on suprasegementals. This course is taught in year 2 of the English Teacher Training Program offered at a local state university in Mar del Plata, Argentina. The students enroll in this class after having taken and passed *English Phonetics and Phonology I* and *II*, where they study the nature of English sounds.

Fifty Spanish-L1 trainees participated in this study. They were divided into two groups: A ($n = 25$) and B ($n = 25$). In Group A, learners were from 20 to 41 years old ($M = 22$), and in Group B, they were from 19 to 30 years old ($M = 22$). Each group consisted of 23 females and 2 males. Their level of English language competence before entering the university was equivalent to a TOEFL iBT Total Score of 70 or above. Their formal L2 instruction ranged from 5 to 9 years at private local language institutes in Mar del Plata, Argentina (M years of instruction: A = 7.12, B = 7.32). None reported having lived in an English-speaking country before taking the course. When data were gathered, these learners were taking other courses in English in the same teacher training program. In both groups, the participants reported not having used English outside the classroom, except for completing homework. Native and non-native English speaker models (their teachers and listening materials) were the type of input they received.

4.2 Teaching Intervention

Both groups were taught during a 16-week period at different times and received suprasegmental instruction using a teacher-centered approach to pronunciation teaching (TCT), which focused on form. The theoretical sessions of instruction lasted 16 weeks and covered aspects relating to English stress, rhythm and intonation. Practical sessions consisted in dictations whereby students were required to recognize and transcribe segmental and prosodic features using phonetic script and pronunciation conventions. Students completed controlled exercises, imitating British English native-speakers using the RP (Received Pronunciation) accent. Student-teacher and student-student interactions were limited. Unlike Group A, Group B included a communicative, awareness-building component (CABC) with a strong focus on the teaching of suprasegmentals. This CABC was taught within the same timeframe as the other group in a weekly 2-h block. To include this component, a lab controlled-practice block was taken out from the instruction. Within the CABC, learners completed a battery of communicative tasks aimed at raising their awareness of specific phonological target forms followed by a period of analysis and reflection (Luchini, 2018). These tasks required students to work collaboratively in class. They were asked to recognize, analyze, reflect and emulate different phonological target forms as well as to self-assess their productions. They were exposed to different native and nonnative English accents.

4.3 L2 Speech Samples

The speech samples were taken from task 2 (T2) of an oral achievement test (Luchini, 2004) administered to participants before and after instruction in the form of pre/post-tests at weeks 1 and 16, respectively. T2 asked learners to compare and contrast two pictures of people doing different activities. This task was chosen because it does not present any interactional phenomena, thus facilitating the data processing as there are no voice overlaps, or changes in tonal adaptation caused by turn taking. Thirty seconds of recording were selected from both the pre- and post-tests which lasted approximately two minutes each. Each speech sample was delimited by the use of two-time markers: after the first 10 s of starting T2, and within the stipulated 30 s. That is, neither the beginning nor the end of the task was included for analysis, thus allowing the study of the central portion of all the recordings, which is the extension of the speech signal with a greater degree of fluency. Working with standardized speech samples in terms of their duration allows for more consistent comparisons among productions.

4.4 Assessment Procedure for Nuclear Stress Placement

Two experienced English pronunciation teacher-researchers (one a balanced Spanish/English bilingual speaker, the other a Spanish-L1 speaker and advanced English-L2 speaker) worked independently, listened to, transcribed, and segmented the spelling transcripts of the students' recordings into tone units. The tone unit is a unit of English phonology, which can be defined as one melodic contour (Halliday & Greaves, 2008). Both filled and empty pauses were removed from the spelling transcripts. The teachers identified and placed nuclear stresses in each tone unit following the rules that govern English nuclear stress, and informed by the context provided by the transcripts. To measure the assessors' degree of agreement, interrater reliability was used. That is, the results of each assessor were compared in order to determine consensus degree. The percentage of homogeneity between results was 89%. Assessors negotiated final agreement for the remaining 11%. The assessors reported that segmenting the orthographic transcripts presented a high degree of complexity, resulting in, for example, one same statement being segmented differently. To standardize this process, both assessors were asked to arrive at a *standard response* against which they would later evaluate the learners' speech samples. Therefore, the assessors compared the learners' recordings to the standard response to determine the correct/incorrect location of nuclear stresses. That is, for each participant's speech sample in both pre/post-test conditions, there was a corresponding standard response (100 speech samples = 100 standard responses). To operationalize this procedure, Trofimovich's personal suggestion was followed (Trofimovich, October, 2009, personal conversation):

> The problem with working with free speech data is normalizing for speech length. Different people produce speech samples that are different in length and therefore they have different opportunities to produce the items that you are measuring. To get a measure for each participant, I would divide each count by the total number of possibilities. These "counts" and "possibilities" depend on what you are measuring. If, for example, in a given speech sample there are five nuclear stresses (as suggested by the specialists), then you have your 5 possibilities. So everything the student has done will be counted out of 5. If, for example, a student got 3 out of 5 of these stresses right, then your correct count for this student is 3. And your final measure will be 3 divided by 5 (3/5), that is, 3 stresses out of 5. So your counts will be a proportion of nuclear stresses produced correctly. If a student got more stresses than there should be, I would not punish him/her for these "extra" stresses.

This procedure consists in dividing the total number of coincidences in the students' productions of nuclear stresses by those agreed in the assessors' standard responses. For example, if for the same speech sample, a learner marked three nuclear stresses while the assessors agreed on five, the average for that participant was 3 out of 5 (3/5), equivalent to: 0.6. Whenever learners' production evidenced a greater number of nuclear stresses than those identified by the assessors, those were not considered as part of the total average. Only nuclear stresses that matched the standard response number counted as correct answers. This procedure is illustrated in Table 1, using Student 9's speech sample in pre-test condition (Group A) as an example. Slanted bars (/) indicate tone unit boundaries. Syllables bearing the nuclear stress are shown in **bold font**.

Whereas the assessors identified six nuclear stresses in the standard response, the analysis of Student 9's speech sample shows 8 nuclear stresses, 3 of which matched the ones they had identified in the standard response. Following Trofimovich's suggestion, then, the value of (1) represents total coincidence of nuclear stress placement in both samples. In this case, it indicates a coefficient of 0.5.

Table 1 Pre-test condition: sample of standard response and assessors' perceptual analysis

Standard response *Pre-test*	Perceptual analysis *Pre-test*
/there's just one young **man**/ I think he might be playing some kind of **ins**trument/ or he has drunk many **so**das apparently/ In the **first** picture/ they have like a **lunch** meeting or something/ probably there are **more** people/	/there's just one young **man**/ I think he might be playing some kind of **ins**trument/ or he has drunk many **so**das apparently/ In the first **pic**ture/ they have like a lunch **meet**ing/ or **some**thing/ **prob**ably/ there are more **peo**ple/

Table 2 Post-test condition: sample of standard response and assessors' perceptual analysis

Standard response *Post-test*	Perceptual analysis *Post-test*
The first **two** teenagers/ probably are at a **table/** having a lot of **food/** I cannot say **what** it is/ or what **kind** of food it is/ but it doesn't look very **health**y/ They are like at the **din**ner party/ or something of the **sort/** And in the **sec**ond photograph/ they seems to be at the **street/**	The first two teen**age**rs probably/ are at a **table/** having a lot of **food/** I cannot **say/** what it is or what kind of **food** it is/ but it doesn't look very **health**y/ They are like at the **dinn**er party/ or something of the **sort/** And in the second **pho**tograph/ they seems to be at the **street/**

The same assessment procedures were used to analyze Student 9's production in the post-test condition (Group A). Table 2 shows both the standard response and the assessors' perceptual analysis.

From their experience as pronunciation teachers and guided by phrase stress rules, the assessors identified ten nuclear stresses which they highlighted in their standard response. Interestingly, in their perceptual analysis of Student 9's production, they also recognized ten nuclear stresses, of which six matched those they had identified in the standard response. The results in the post-test condition reveal a coefficient of 0.6, indicating an improvement in the location of nuclear stresses.

Two acoustic profiles of the same phrase "*and in the second photograph*" taken from this same speech sample are shown below. The computer program *Speech Analyzer*® for acoustic analysis of speech sounds was used (see https://software.sil.org/speech-analyzer/). Figure 1 shows Student 9's production taken from the post-test condition. Figure 2 shows the production of the same phrase coming from the balanced Spanish/English bilingual assessor. The student misplaces the nuclear stress on the word "*photograph*," while the assessor, in her role as bilingual speaker and specialist correctly locates the stress on the word "*second*," consistent with the rules that govern contrastive stress placement in English.

Figure 1 shows three acoustic records in four windows that allow visualization of the variations in the speech wave and intensity (on the left), and frequency (F0) and harmonics (on the right). All reflect the physical properties with which Student 9 emitted the phrase "*and in the second photograph*." In the windows located on the right-hand side, for example, two important peaks of F0 are located: the first of approximately 200 *Hz* corresponding to "*second*," and the most significant is 258 *Hz* in '*photograph*,' which acquires all the requirements of a nuclear stress, peak intensity of around 60 dB (decibels) being added to these F0 values.

Figure 2 illustrates the extent of individual variations in the production of prosodic contours. In principle, and always within the comparison of the same phrase emitted by two female speakers, it is observed that nuclear stress assignation is different. In this case, the assessor decides to highlight the first syllable of the word "*second*," because she has taken into account the variants of the discursive context, always bearing in mind the order of the implementation of T2. The need to use contrastive

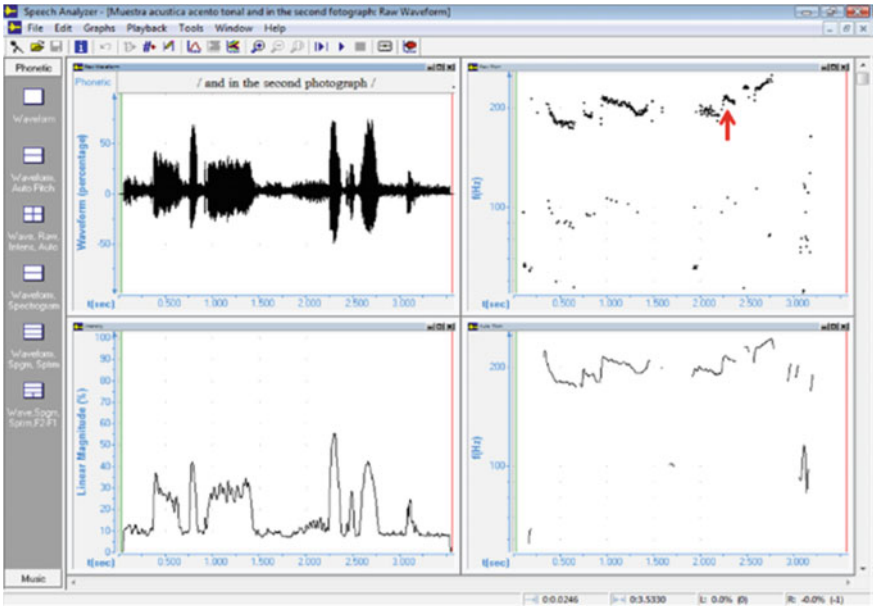

Fig. 1 Acoustic profile of the phrase "*and in the second* ***photo****graph*" issued by the student (the arrow indicates nuclear stress location)

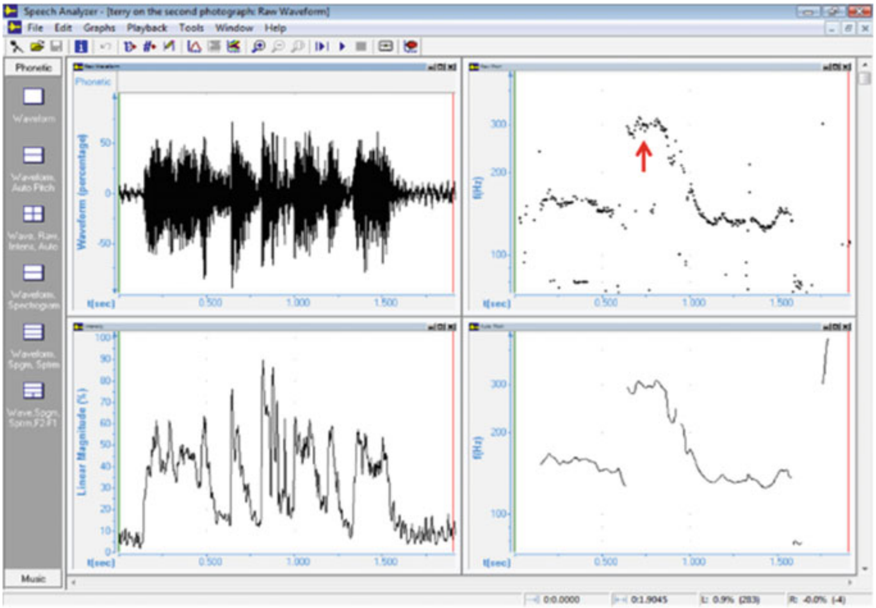

Fig. 2 Acoustic profile of the phrase "*and in the* ***sec****ond photograph*" issued by the assessor (the arrow indicates nuclear stress location)

stress in this case leads her to produce a 310 Hz frequency peak that accompanies an energy rise of around 89 dB. To her sharp register, we must add what is displayed in the first window on the left: the fluency of her continuous speech flow, which does not present any sign of hesitation, thus there is no explicit pause. This is clearly shown by the compression of the waveform, where we can see a noteworthy difference between the two speakers' realizations of the same phrase. In Fig. 1, this same window presents hesitation or doubt, equivalent to a pause between the stresses mentioned (see distance in the temporal axis of the windows located to the right of Fig. 1, between the words "*second*" and "*photograph*").

Observation of these acoustic profiles allows us to distinguish the scope of individual differences and degrees of training in the use of English pronunciation between the two speakers. The acoustic aspect that is associated with a better command of English is manifested by the balanced bilingual speaker. This type of acoustic analysis highlights the importance of the use and localization of the nuclear stress. We can see that the operationalization of this assessment procedure provides interesting information that allows to evaluate the efficacy of speakers' use and placement of nuclear stress in free speech production in English and to measure their linguistic development and oral proficiency over time.

4.5 Assessment Procedure for Comprehensibility

Ten experienced English native speaker raters, operating independently, listened to the 100 speech samples and rated them using a Likert-like scale ranging from 1–9 to determine the speakers' degree of perceived comprehensibility. In this context, comprehensibility means the perceived ease or difficulty of understanding L2 speech, that is, the cognitive effort made by the listener to understand accented speech (Munro & Derwing, 1995). The listeners made a scalar judgment of comprehensibility where 9 indicated total ease of understanding and 1 showed poor comprehensibility or high degree of effort on the part of the listener to understand non-native speech. The raters heard each stimulus once. To reduce the effects of fatigue, they were given short breaks in between the recordings. None of the raters reported having had hearing impediments.

4.6 Data Analysis

The statistical analysis was performed using *InfoStat* software v2020e (Di Rienzo et al., 2014). Mean values, standard deviation and least significant differences were calculated using the *LeastSquares Fit model*. Mean comparisons were made with independent samples *t*-tests in each group for both independent variables (nuclear stress and comprehensibility) with $\alpha = 0.05$. A simple linear regression analysis was performed between the relative responses (RR) of nuclear stress and comprehensibility: RR = (post-test − pre-test / pre-test) × (100) to determine whether the variations in one of the variables explain the variations in the other. Outliers were removed prior to analysis.

5 Results

In either treatment, no significant differences were observed between pre- and post-test conditions for both nuclear stress and comprehensibility (see Tables 3 and 4).

However, a slight tendency for improvement for nuclear stress and higher values for comprehensibility are observed between pre- and post-tests in the CABC treatment, which indicates that Group B improved the nuclear stress coincidence coefficient (i.e., closer to 0.8) and the degree of comprehensibility (i.e., higher than 6) (see Figs. 3 and 4).

The simple linear regression analysis between RRNS (relative response nuclear stress) and RRC (relative response comprehensibility) for Group A (TCT) was not statistically significant with a *p-value* of 0.6073 and an *adjusted R-squared* of −0.03421. On the other hand, the simple linear regression analysis of Group B (CABC) was statistically significant with $\alpha = 0.1$ (probability error of 10%) with a *p-value* of 0.0542 and an *adjusted R-squared* of 0.1383.

Figure 5 shows the dispersion of the data between the relative responses of the analyzed variables (RRNS and RRC), and the simple positive linear regression line is shown only for Group B. The improvements obtained by this group, exposed to the CABC treatment, explain 14% of the variation in RRC, while Group A, under the TCT treatment, does not register a relationship between the variables. This indicates the lack of relationship between the variation in nuclear stress placement throughout the instructional period and the variations in comprehensibility.

6 Discussion

Both experimental groups underwent a teacher-centered treatment (TCT), but only Group B included a communicative-awareness-building component (CABC). Intra-group analyses revealed that Group B obtained better results in the two variables analyzed (nuclear stress and comprehensibility), thereby demonstrating that formal instruction that includes a communicative component tends to show better improvements on students' productions. In this communicative block, Group B learners completed a battery of progressive tasks aimed at raising their awareness of specific phonological target forms followed by a period of analysis and reflection. These tasks were sequenced in order to lay emphasis on a meaning-form-meaning progression that sought to recognize phonological gaps in the students' interlanguage while in the process of constructing meaning. Phonological gaps were

Table 3 Intragroup mean comparison for nuclear stress (independent samples *t*-tests)

Treatment	G1	G2	*M* (1)	*M* (2)	$M(1) - M(2)$	LI (95)	LS (95)	pHomVar	*t*	*p*	Test
CABC (n = 21)	(post-test)	(pre-test)	0.80	0.72	0.08	−0.03	0.18	0.3190	1.46	0.1518	Bilateral
TCT (n = 23)	(post-test)	(pre-test)	0.78	0.78	−4.8E-03	−0.010	0.09	0.8308	−0.11	0.9156	Bilateral

Table 4 Intragroup mean comparison for comprehensibility (independent samples *t*-tests)

Treatment	G1	G2	*M* (1)	*M* (2)	*M* (1) − *M*(2)	LI (95)	LS (95)	pHomVar	*t*	*p*	Test
CABC (n = 21)	(post-test)	(pre-test)	6.19	6.11	0.08	−0.31	0.46	0.9560	0.40	0.6892	Bilateral
TCT (n = 23)	(post-test)	(pre-test)	6.07	5.96	0.10	−0.53	0.74	0.8393	0.33	0.7425	Bilateral

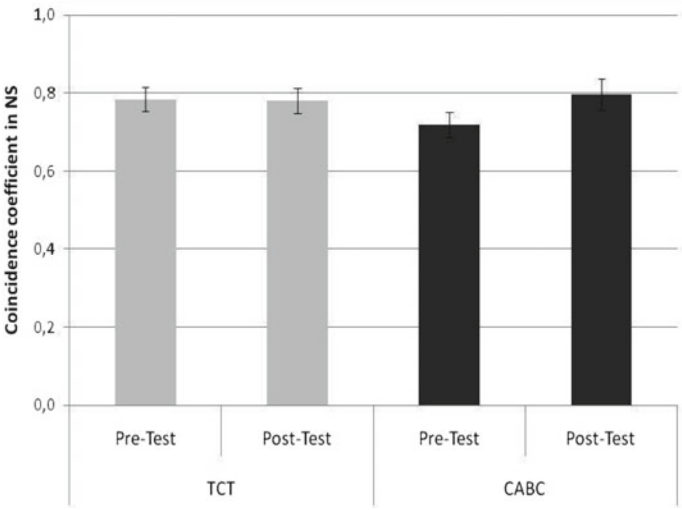

Fig. 3 Coincidence coefficient in nuclear stress

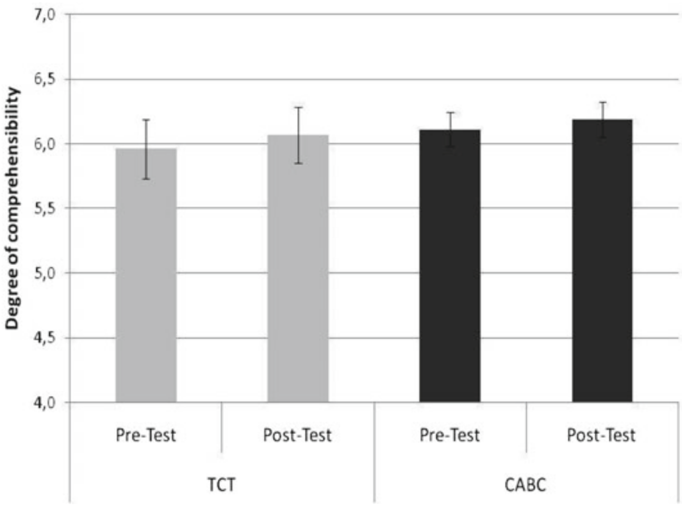

Fig. 4 Degree of comprehensibility

filled in language-pronunciation focused sessions by making comparisons between input and output which brought about discussions, always focusing on phonological target forms (Samuda, 2001; Swain & Lapkin, 2001). The aim of these tasks was to help students raise their awareness of key phonological features and the

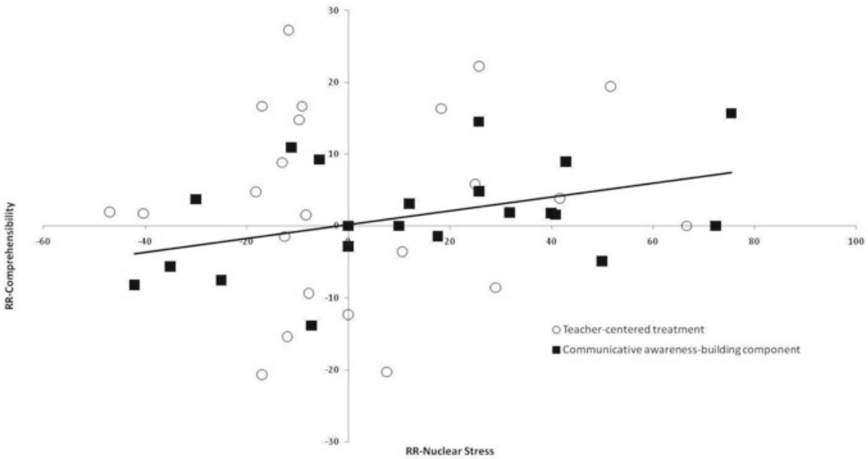

Fig. 5 Simple linear regression analysis between RRNS and RRC for TCT (Group A) and CABC (Group B) treatments

contribution of these aspects for establishing both receptive and productive intelligibility/comprehensibility. The inclusion of CABC promoted some degree of improvement in the target form, and thus provides the groundwork for more accurate nuclear stress placement.

The simple linear regression analysis between RRNS and RRC in Group A was not statistically significant, showing no relationship between the variables. The pedagogical treatment applied to Group A (TCT) was more aligned with the *Nativeness* paradigm. This finding echoes findings from other studies framed within the *Nativeness* paradigm (see Thomson & Derwing, 2014), which showed an unclear relationship between pronunciation improvement with discrete features and more intelligible and comprehensible speech. In contrast, a statistically significant linear regression between RRNS and RRC was found for Group B, which included the CABC treatment (more aligned with the *Intelligibility* principle). Group B's 14% improvement in comprehensibility was due to students' progress in nuclear stress placement. Thus, the results of the current study are more in line with previous findings showing that the *Intelligibility*-oriented paradigm is highly related to better achievements in intelligibility and comprehensibility (Isaacs & Trofimovich, 2012; Jułkowska & Cebrian, 2015; Kang, 2010; Saito et al., 2015, 2016; Thomson, 2018; Trofimovich & Isaacs, 2012). Isaacs and Trofimovich (2012) amalgamated intelligibility and comprehensibility suggesting that issues concerning comprehensibility are consistent with the instructional goals of helping learners attain intelligible pronunciation. It appears then that an awareness-rising, communicative-based approach for the teaching of suprasegmentals improves nuclear stress placement which, in turn, brings about advances in comprehensibility. Therefore, it seems to be advisable for pronunciation teachers to incorporate form-focused classroom tasks that strictly aim to develop the correct use and localization of nuclear stress enabling students to achieve better comprehensibility. The CABC instructional block was limited to two weekly hours. The question for further investigation remains: Had the students received more instructional hours within the CABC framework, would their degree of comprehensibility have been better?

The data from this study reinforce and complement the pedagogical claim that emphasizes the importance of teaching nuclear stress placement to achieve higher levels of comprehensibility (Hahn, 2004; Kang, 2010; Morley, 1991). In her Lingua Franca Core, Jenkins (2000, 2002) includes nuclear stress as an essential feature because it contributes to mutual intelligibility in interactions among non-native speakers from different linguistic backgrounds. Along these lines, McNerney and Mendelsohn (1992) affirm that suprasegmentals must be taught before any other phonological aspect because they have a direct impact on the students' comprehensibility. They argue that giving priority to suprasegmentals not only improves students' comprehensibility but also contributes to raising their self-esteem as greater changes can be effected in their speech. Similar pedagogical arguments have been put forward by some other researchers (Brazil et al., 1980; Brown, 1995; Celce-Murcia et al., 2010; Clennell, 1996; Derwing et al., 1998; Kang, 2010; Kang & Johnson, 2018; Morley, 1991; Pennington & Ellis, 2000; Pennington & Richards, 1986). These findings lead to important pedagogical implications for language teachers, program designers and developers. Pronunciation classes can focus—though not exclusively—on suprasegmental differences directly related to nuclear stress placement and listeners' comprehensibility.

Finally, the design and systematization of an assessment protocol for measuring L2 pronunciation features permeates this chapter. The speech measures were analyzed using both auditory and instrumental techniques. Ten raters measured comprehensibility using a numerical scale of listener perception. For nuclear stress measurements, two specialized teacher-researchers provided a standard response for each speech sample that was later compared against the students' productions. Nuclear stress measurements were then compared with those of comprehensibility using simple linear regression analysis. This comparison determined the efficacy of one treatment over the other. The main findings of this study draw light on the usefulness of the implemented pronunciation assessment protocol for evaluating the effectiveness of L2 pronunciation instruction. This protocol may be valuable for other teachers and researchers interested in measuring students' L2 pronunciation gains and compare results in other teaching contexts using the same or other pedagogical treatments.

6.1 Limitations

Although the findings of this study yielded interesting results regarding the impact of one L2 pronunciation pedagogical treatment over another, there are a number of limitations that are worth discussing. As opposed to lab experiments, classroom-based studies, such as this one, inevitably allow for other influencing linguistic factors that need to be taken into account. The students in both groups were simultaneously taking other English classes along with DOII. This classroom input could have enhanced their L2 pronunciation development as well. Both groups showed a slight tendency towards improvement in their oral productions after instruction. Yet, Group B scored higher results than Group A. This intragroup difference is attributed to the inclusion of the CABC in the pedagogical treatment applied to Group B. As already mentioned, it is worth posing the question whether a larger CABC workload would have produced further advancement in comprehensibility. Moreover, a comparative study that evaluates the impact of two entirely different treatments, one of them being exclusively based on the CABC pedagogical principles, would be needed.

The current study involved the measurement of one single prosodic feature: nuclear stress placement. Research on measurements of other elements of oral proficiency such as lexical stress, rhythm and pause duration, location and frequency, for example, would have yielded more information to corroborate or contradict the findings obtained. Additionally, learning about the students' beliefs and perceptions regarding the development of their L2 productions before and after instruction would have allowed to cross-check different types of data, confirm findings, and perhaps help to interpret the results obtained. Unfortunately, due to time constraints, neither option was possible. Finally, all the participants were L1-Spanish speakers studying to become English teachers. Results from populations of other cultural and linguistic backgrounds, and learning goals could provide other types of information.

7 Implications

This study may be a valuable contribution to future teachers and researchers. Correct nuclear stress placement contributes to increased comprehensibility and as such should be given high priority in the L2 pronunciation class. These findings may also lead to important pedagogical implications, providing additional evidence to support the importance of suprasegementals on listeners' perception and clarify their relationship with comprehensibility on L2 learners' speech. Teachers, trainers, program designers and developers should focus their attention on devoting more class time to teach this prosodic feature.

Furthermore, a pedagogical proposal for the teaching of suprasegmentals, as the one deployed in the CABC group, whereby tasks function as a pivot to develop phonological awareness, promotes self-efficacy and pushes learners to use metacognitive skills. Both these factors increase the chances of further development of comprehensible read-aloud L2 oral production (Sardegna, 2012, 2021). The current study contributes to this line of research by showing how improvements in nuclear stress placement via awareness-raising relate to improvements in speech comprehensibility in a picture description task. Further research in this area is needed to confirm and extend the findings to other speech features.

8 Conclusion

Pronunciation instruction is not a marginalized area of second language teaching and research anymore. This field is growing swiftly. Explicit pronunciation instruction can have a significant effect on students' oral production because it focuses on learners' attention to phonetic information and increases phonological awareness, which promotes learning in a way that exposure alone does not. Previous research on L2 pronunciation has concentrated on identifying L2 learners' stress patterns. Only a few studies have measured the effect of nuclear stress placement along with comprehensibility ratings, and have used this assessment procedure as an instrument to evaluate the effectiveness of L2 pronunciation instruction. The innovative assessment procedure used in this classroom-based study proved useful to determine such efficacy. As mentioned earlier, further studies are needed to corroborate or contradict these findings, as well as to increase our understanding of suprasegmentals as predictors of L2 pronunciation development, and to help define instructional priorities.

Acknowledgements Thanks to Mariela Xynos and Ubiratã Kickhöfel Alves for their comments and suggestions on earlier drafts.

References

Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning, 42*(4), 529–555. https://doi.org/10.1111/j.1467-1770.1992.tb01043.x

Bardovi-Harlig, K. (1986). *Pragmatic determinants of English sentence stress.* Indiana University Linguistics Club.

Birch, S., & Clifton, C. (1995). Focus, accent, and argument structure: Effects on language comprehension. *Language and Speech, 38*(4), 365–391. https://doi.org/10.1177/002383099503800403

Bøhn, H., & Hansen, T. (2017). Assessing pronunciation in an EFL context: Teachers' orientations towards nativeness and intelligibility. *Language Assessment Quarterly, 14*(1), 54–68. https://doi.org/10.1080/15434303.2016.1256407

Brazil, D., Coulthard, M., & Johns, C. (1980). *Discourse intonation and language teaching.* Longman.

Brown, A. (1995). Minimal pairs: Minimal importance? *English Language Teaching Journal, 49*(2), 169–175. https://doi.org/10.1093/elt/49.2.169

Celce-Murcia, M., Brinton, D. M., Goodwin, J. M., & Griner, B. (2010). *Teaching pronunciation: A course book and reference guide* (2nd ed.). Cambridge University Press.

Chun, C. W. (2006). An analysis of a language test for employment: The authenticity of the PhonePass test. *Language Assessment Quarterly, 3*(3), 295–306. https://doi.org/10.1207/s15434311laq0303_4

Chun, C. W. (2008). Comments on 'evaluation of the usefulness of the Versant for English test: A response': The author responds. *Language Assessment Quarterly, 5*(2), 168–172. https://doi.org/10.1080/15434300801934751

Clennell, C. (1996). Promoting the role of English prosody in a discourse-based approach to oral interaction. *Prospect, 11*(3), 17–28.

Dauer, R. (2005). The Lingua Franca Core. A new model for pronunciation instruction? *TESOL Quarterly, 39*(3), 543–549. https://doi.org/10.2307/3588494

Derwing, T., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition, 19*(1), 1–16. https://doi.org/10.1017/S0272263197001010

Derwing, T., & Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching, 42*(4), 1–15. https://doi.org/10.1017/S026144480800551X

Derwing, T., & Munro, M. J. (2015). The interface of teaching and research: What type of pronunciation instruction should L2 learners expect? In P. Luchini, M. A. García Jurado, & U. Alves (Eds.), *Fonética y fonología: Articulación entre enseñanza e investigación* (pp. 14–26). Biblioteca Central Universidad Nacional de Mar del Plata.

Derwing, T., Munro, M., Foote, J., Waugh, E., & Fleming, J. (2014). Open the window on comprehensible pronunciation after 19 years: A workplace training study. *Language Learning, 64*(3), 526–548. https://doi.org/10.1111/lang.12053

Derwing, T., Munro, M., & Wiebe, G. (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

Di Rienzo J., Casanoves, F., Balzarini, M., Gonzalez, L., Tablada, M., & Robledo, C. (2014). *InfoStat* (versión 2014) [Computer software]. Universidad Nacional de Córdoba. https://www.infostat.com.ar/

Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly, 39*(3), 399–423. https://doi.org/10.2307/3588487

Foote, J. A., Holtby, A. K., & Derwing, T. M. (2011). Survey of the teaching of pronunciation in adult ESL programs in Canada, 2010. *TESL Canada Journal, 29*(1), 1–22. https://doi.org/10.18806/tesl.v29i1.1086

Fulcher, G. (2015). Assessing second language speaking. *Language Teaching, 48*(2), 198–216. https://doi.org/10.1017/S0261444814000391

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* (pp. 194–206). Iowa State University.

Hahn, L. D. (2004). Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly, 38*(2), 201–203. https://doi.org/10.2307/3588378

Halliday, M. A. K. (2013). *Halliday's introduction to functional grammar* (4th ed.). Routledge. https://doi.org/10.4324/9780203431269

Halliday, M. A. K., & Greaves, W. (2008). *Intonation in the grammar of English.* Equinox. https://www.equinoxpub.com/home/intonation-grammar-english-m-k-halliday-william-greaves/

Hymes, D. H. (1972). On communicative competence. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics: Selected readings* (pp. 269–293). Penguin.

Isaacs, T. (2008). Towards defining a valid assessment criterion of pronunciation proficiency in non-native English-speaking graduate students. *Canadian Modern Language Review, 64*(4), 555–580. https://doi.org/10.3138/cmlr.64.4.555

Isaacs, T. (2016). Assessing speaking. In D. Tsagari & J. Banerjee (Eds.), *Handbook of second language assessment* (pp. 131–146). De Gruyter Mouton. https://doi.org/10.1515/9781614513827-011

Isaacs, T., & Trofimovich, P. (2012). Deconstructing comprehensibility: Identifying the linguistic influences on listeners' L2 comprehensibility ratings. *Studies in Second Language Acquisition, 34*(3), 475–505. https://doi.org/10.1017/S0272263112000150

Isaacs, T., & Trofimovich, P. (2016). *Second language pronunciation assessment: Interdisciplinary perspectives.* Multilingual Matters. https://doi.org/10.21832/ISAACS6848

Jenkins, J. (2000). *The phonology of English as an international language.* 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

Jułkowska, I. A., & Cebrian, J. (2015). Effects of listener factors and stimulus properties on the intelligibility, comprehensibility and accentedness of L2 speech. *Journal of Second Language Pronunciation, 1*(2), 211–237. https://doi.org/10.1075/jslp.1.2.04jul

Kang, O. (2010). Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System, 38*(2), 301–315. https://doi.org/10.1016/j.system.2010.01.005

Kang, O. (2012). Relative impact of pronunciation features on ratings of non-native speakers' oral proficiency. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference* (pp. 10–15). Iowa State University. https://apling.engl.iastate.edu/wp-content/uploads/sites/221/2015/05/PSLLT_4th_Proceedings_2012.pdf

Kang, O., & Johnson, D. (2018). The roles of suprasegmental features in predicting English oral proficiency with an automated system. *Language Assessment Quarterly, 15*(2), 150–168. https://doi.org/10.1080/15434303.2018.1451531

Kang, O., & Pickering, L. (2014). Using acoustic and temporal analysis for assessing speaking. In A. J. Kunnan (Ed.), *The companion to language assessment* (Vol. 2, pp. 1047–1062). Wiley-Blackwell. https://doi.org/10.1002/9781118411360.wbcla056

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

Kim, H. J. (2015). A qualitative analysis of rater behavior on an L2 speaking assessment. *Language Assessment Quarterly, 12*(3), 239–261. https://doi.org/10.1080/15434303.2015.1049353

Lado, R. (1961). *Language testing: The construction and use of foreign language tests*. Longman.

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. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly, 39*(3), 369–377. https://doi.org/10.2307/3588485

Levis, J. (2018). Intelligibility, oral communication, and the teaching of pronunciation. *Cambridge University Press*. https://doi.org/10.1017/9781108241564

Luchini, P. (2004). Designing a pronunciation test for assessing free speech production: An evaluative case study. *IATEFL Speak Out!, 31*, 12–24. https://pronsig.iatefl.org/journal/

Luchini, P. (2017). Measurement for accentedness, pause frequency/duration and nuclear stress placement in the EFL classroom. *Ilha Do Desterro, 70*(3), 185–200. https://doi.org/10.5007/2175-8026.2017v70n3p185

Luchini, P. (2018). *Suprasegmental phonology: Handbook for pronunciation skill teaching. Integrating theory with practice* (2nd ed.). Biblioteca Central Universidad Nacional de Mar del Plata. http://biblio1.mdp.edu.ar/centro-ventas/producto/suprasegmental-phonology-pronunciation-skill-teaching/

McNerney, M., & Mendelsohn, D. (1992). Suprasegmentals in the pronunciation class: Setting priorities. In P. Avery & S. Ehrlich (Eds.), *Teaching American English pronunciation* (pp. 185–196). Oxford University Press. https://elt.oup.com/catalogue/items/global/teacher_development/oxford_handbooks_for_language_teachers/9780194328159?cc=global&selLanguage=en

Morley, J. (1991). The pronunciation component of teaching English to speakers of other languages. *TESOL Quarterly, 25*(3), 481–520. https://doi.org/10.2307/3586981

Munro, M. J. (1995). Nonsegmental factors in foreign accent: Ratings of filtered speech. *Studies in Second Language Acquisition, 17*(1), 17–34. https://doi.org/10.1017/S0272263100013735

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

Murphy, J. (2014). Intelligible, comprehensible, non-native models in ESL/EFL pronunciation teaching. *System, 42*(1), 258–269. https://doi.org/10.1016/j.system.2013.12.007

Pennington, M., & Ellis, N. (2000). Cantonese speakers' memory for English sentences with prosodic cues. *The Modern Language Journal, 84*(3), 372–389. https://doi.org/10.1111/0026-7902.00075

Pennington, M., & Richards, J. (1986). Pronunciation revisited. *TESOL Quarterly, 20*(2), 207–225. https://doi.org/10.2307/3586541

Pennington, M., & Rogerson-Revell, P. (2019). *English pronunciation teaching and research.* Palgrave MacMillan. https://doi.org/10.1057/978-1-137-47677-7

Saito, K. (2011). Examining the role of explicit phonetic instruction in native-like and comprehensible pronunciation development: An instructed SLA approach to L2 phonology. *Language Awareness, 20*(1), 45–59. https://doi.org/10.1080/09658416.2010.540326

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., Trofimovich, P., & Isaacs, T. (2015). Using listener judgments to investigate linguistic influences on L2 comprehensibility and accentedness: A validation and generalization study. *Applied Linguistics, 38*(4), 1–25. https://doi.org/10.1093/applin/amv047

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, Y., & Saito, K. (2017). 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

Samuda, V. (2001). Guiding relationships between form and meaning during task performance: The role of the teacher. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and testing* (pp. 119–140). Longman. https://doi.org/10.1075/tblt.1.20gui

Sardegna, V. G. (2012). Learner differences in strategy use, self-efficacy beliefs, and pronunciation improvement. In. J. Levis & K. LeVelle (Eds.), *Proceedings of the 3rd Pronunciation in Second Language Learning and Teaching Conference* (pp. 39–53). Iowa State University.

Sardegna, V. G. (2020). Pronunciation and good language teachers. In C. Griffiths & Z. Tajeddin (Eds.), *Lessons from good language teachers* (pp. 232–245). Cambridge University Press. https://doi.org/10.1017/9781108774390.021

Sardegna, V. G. (2021). Evidence in favor of a strategy-based model for English pronunciation instruction. *Language Teaching,* 1–16. Advance online publication. https://doi.org/10.1017/S0261444821000380

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition.* Harvard University Press.

Swain, M., & Lapkin, S. (2001). Focus on form through collaborative dialogue: Exploring task effects. In M. Bygate, P. Skehan & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and testing* (pp. 99–118). Longman. https://doi.org/10.4324/9781315838267-14

Tajima, K., Port, R., & Dalby, J. (1997). Effects of temporal correction on intelligibility of foreign-accented English. *Journal of Phonetics, 25*(1), 1–24. https://doi.org/10.1006/JPHO.1996.0031

Terken, J., & Hirschberg, J. (1994). Deaccentuation of words representing "given" information: Effects of persistence of grammatical function and surface position. *Language and Speech, 37*(2), 125–145. https://doi.org/10.1177/002383099403700202

Thomson, R. (2018). Measurement of accentedness, intelligibility and comprehensibility. In O. Kang, & A. Ginther (Eds.), *Assessment in second language pronunciation* (pp. 11–29). Routledge. https://doi.org/10.4324/9781315170756-2

Thomson, R., & Derwing, T. (2014). The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics, 36*(3), 326–344. https://doi.org/10.1093/applin/amu076

Trofimovich, P., & Isaacs, T. (2012). Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition, 15*(4), 905–916. https://doi.org/10.1017/S1366728912000168

Trofimovich, P., Isaacs, T., Kennedy, S., Saito, K., & Crowther, D. (2016). Flawed self-assessment: Investigating self- and other-perception of second language speech. *Bilingualism: Language and Cognition, 19*(1), 122–140. https://doi.org/10.1017/S1366728914000832

Trofimovich, P., Kennedy, S., & Blanchet, J. (2017). Development of second language French oral skills in an instructed setting: A focus on speech ratings. *Canadian Journal of Applied Linguistics, 20*(2), 32–50. https://doi.org/10.7202/1042675AR

Winters, S., & O'Brien, M. G. (2013). Perceived accentedness and intelligibility: The relative contributions of F0 and duration. *Speech Communication, 55*(3), 486–507. https://doi.org/10.1016/j.specom.2012.12.006

Xi, X. (2010). Automated scoring and feedback systems for language assessment and learning. *Special Issue of Language Testing, 27*(3), 291–300. https://doi.org/10.1177/0265532210364643

Xi, X. (2012). Validity and the automated scoring of performance tests. In G. Fulcher & F. Davidson (Eds.), *The Routledge handbook of language testing* (pp. 438–451). Routledge. https://doi.org/10.4324/9780203181287

Pedro Luis Luchini is Full Professor and Research Group Director for "Cuestiones del Lenguaje" at Universidad Nacional de Mar del Plata, Argentina. Additionally, he participated in a Fulbright Exchange Program at College of DuPage, Illinois, US (1997–1998) and a Faculty Enrichment Program at Concordia University (CU), Montreal, Canada (2007), and taught EFL at Shanghai Normal University, China (2003–2004). He received a doctoral research award at CU in 2009. His research mainly addresses issues on Applied Linguistics with a focus on L2 English pronunciation.

Cosme Daniel Paz is a Ph.D. student and graduate assistant in Agricultural Sciences, Universidad Nacional de Mar del Plata (UNMdP), Argentina, and Member of the Research Group "Cuestiones del Lenguaje" at UNMdP-ANPCYT-INTA. He graduated with an agricultural engineering degree from Universidad Nacional de Salta (2011), and was awarded a CONICET doctoral scholarship (2012–2016). He is particularly interested in performing statistical analysis related to L2 speech development.

The Effects of Prestige Model Familiarity on Students' Perceptions of and Interactions with Diverse English Accents

Gemma Archer

Abstract For decades, English language students worldwide have been exposed to prestige L1 accent models, such as Received Pronunciation (RP) and General American (GA), often to the exclusion of all other varieties. However, the dominance of these models is questionable today with L2-L2 English communication considered the most common interaction pattern globally. This study was undertaken to understand the extent to which the exposure of L2 students to these models can affect their perception of diverse Englishes, such as those encountered while attending an international university. Thirty international students completed a pre-test questionnaire about their accent beliefs, followed by a Verbal Guise test. Although the questionnaire affirmed broadly positive opinions of prestige models, L2 accents were believed to be easier to understand, while the local model, Scottish Standard English (SSE), was believed to be extremely difficult. Results from the Verbal Guise test indicated most positive associations with the RP accent. However, participants selected the L2 Chinese accent as that which they liked the most, just surpassing the prestige models, potentially due to its native-like speech rate. Interestingly, though the local accent, SSE, was perceived poorly in the pre-test questionnaire, the Verbal Guise test results revealed far less severe views.

Keywords Accent · Perception · Prestige models · RP · Global Englishes · Regional Englishes

G. Archer (✉)
University of Strathclyde, Glasgow, Scotland
e-mail: gemma.archer@strath.ac.uk

V. G. Sardegna and A. Jarosz (eds.), *Theoretical and Practical Developments in English Speech Assessment, Research, and Training*, Second Language Learning and Teaching,
https://doi.org/10.1007/978-3-030-98218-8_5

1 Introduction

English is spoken as a second language (L2) with increasing prevalence worldwide, with current estimations of two billion plus users (Crystal, 2019; The British Council, 2013), which far outnumbers native or first language (L1) users. Due to such demographics, it is highly likely that L2-L2 and L2-L1 contact will occur more frequently (Baese-Berk et al., 2013), especially in international universities where the number of English medium educational courses continues to grow yearly (Myhovych, 2019). This is certainly the case in UK universities, where students and academic staff interact using diverse L2 and L1 regional accents. However, accent is a known variable that can affect comprehension significantly (Buck, 2001). As such, it can be startling for many international students to begin a degree programme in a location where the accents they hear do not match the controlled prestige native speaker or 'default' (Sung, 2016) models they were exposed to as they prepared for international university education in their home countries.

The University of Strathclyde in Glasgow, Scotland, where this study occurred, accurately demonstrates the sheer variety of Englishes that can be present within just one institution. University staff represent 77 different nationalities and, in some departments (e.g., University of Strathclyde's Engineering Department, n.d.), almost 50% of the academic staff are international. The university also has a large international student body with their respective mother tongues and accents, and local Scottish staff and students who speak Scottish Standard English (SSE). SSE largely mirrors the written standard of English but, as would be expected, it has its own phonological inventory made up of 13 vowel sounds and 25 consonant sounds. Outside of the university, international students may also encounter the Scots language with its own unique vocabulary, syntax, and phonology. This adds yet another element into the linguistic melting pot with which international L2 students have to contend, often with no warning or preparation, when they study in Glasgow.

Locations such as Strathclyde are not unique, however. In many educational institutions in the UK and beyond, international L2 speakers must grapple with spoken English in its many variations on a daily basis as reported in studies conducted in Singapore (Goh, 1999), the US (Major et al., 2005), Hong Kong (Sung, 2016), and Australia (Harding, 2008). Yet, despite the increasing internationalisation of universities, the spread of English in lingua franca contexts and the further intermingling of L1 and L2 Englishes, literature investigating L2 speakers' perceptions of diverse accents remains limited and classroom materials using prestige native speaker models persist. The goal of this small-scale study is to contribute to this existing body of work by investigating L2 speakers' perceptions of various English accents, many of which international students at the University of Strathclyde are not likely to have had previous exposure to.

2 Literature Review

2.1 *Prestige Accent Models in ELT*

Many variables can influence how comprehensible a speaker is perceived to be, but accent is often held responsible for causing communication breakdowns. However, despite the receptive benefits which could come from increased access to and familiarisation with diverse English accents, the majority of English teaching coursebooks and materials continue to present only two models: the prestige L1 varieties of Received Pronunciation (RP) and General American (GA) (Kiczkowiak, 2021).

When discussing the continued use of the prestige model RP in English language teaching, Jenkins (2007) concluded that feelings towards this model as a standard are "historically deep rooted and thoroughly naturalised" (p. 33). RP is also suggested to be "the most popular accent for EFL purposes … throughout the twentieth century" (Przedlacka, 2008, p. 18) and "the most thoroughly described accent of English" (Wells, 1982, p. 279). A large number of student textbooks, audio materials, teacher training courses, dictionaries and influential phonetics textbooks continue to take Daniel Jones' original descriptions of RP as their pronunciation model de rigueur. In fact, some coursebook writers have even revealed that they face pressure from publishing houses to use only educated L1 speakers from the south east of England in the audio resources they create (Kiczkowiak, 2021).

A second prestige model, GA, also referred to as Standard American English, has become increasingly visible since World War II due to the status of the US as a political and economic power (Crystal, 2019). In addition, the prevalence of the US (and also GA) in films, television, music and online has contributed to an instantly accessible "American cultural hegemony" worldwide (Henderson et al., 2012, p. 21). Considering the growing ease of access to GA and in multiple different formats, it is easy to see why it is considered a viable alternative to RP within ELT and is often the preferred model for many students (Henderson et al., 2012).

2.2 *Factors Affecting Perception of Diverse Accents*

As communication between L2 English speakers of diverse language backgrounds grows, more research investigating L2 interactions and perceptions of accented English is increasingly warranted (Crowther et al., 2016). However, when undertaking such investigations, it is necessary to establish what influences listeners' judgements, be it "the acoustic and phonological properties, or whether they indicate something about the listener and therefore vary with listeners' language experience" (Witteman et al., 2013, p. 537).

In their comprehensive article on non-native listeners' perceptions of accent, Crowther et al. (2016, p. 161) summarise the phonological features which can increase ratings of accentedness and, consequently, the assigning of the label 'native' or 'non-native' speaker. The items listed include accurate production of segmental and suprasegmental features, such as vowel and consonant sounds, syllable stress, syllable length, and pitch. The consequences or 'costs' for the listener upon being confronted with speech realisations that are unfamiliar to them can be an increase in the length of time required for cognitive processing (Adank & McQueen, 2007; Adank et al., 2009; Harding, 2008; Perry et al., 2017) and the obvious potential for misunderstandings (Munro & Derwing, 1995). It is possible that the longer it takes to identify what a speaker has said, the higher the chances a listener will perceive the accent as different or challenging.

Speech rate has also been identified as another potential factor of influence (Trofimovich & Baker, 2006). Speech rate can aid perceptions of native or non-native status and level of proficiency. Native speakers and more proficient users typically receive a higher rate; the opposite occurs among L2 learners and those of lower proficiency (Munro, 1999). Additionally, while qualitative data from L2 listener-raters show they perceive speech rate, particularly faster speech, to be a negative factor reducing their comprehension (Goh, 1999; Harding, 2008; Zhao, 1997), quantitative data have produced mixed and sometimes contradictory results (Munro & Derwing, 1998; Zhao, 1997).

A further variable reported in previous research is that of a shared L1. Studies such as Bent and Bradlow (2003) have shown that a shared L1 between speaker and listener can positively impact on the intelligibility ratings given by the listener, even if the speaker is of lower proficiency. This is due to a phenomenon known as the *Matched Interlanguage Speech Intelligibility Benefit* (Bent & Bradlow, 2003). Kang et al. (2016) succinctly explain this benefit as a feature "which predicts that a NNS listener may be better equipped to interpret specific acoustic–phonetic features of an L2 that are matched with his own L1 than a different L1" (p. 2). However, studies by Tauroza and Luk (1997) found limited evidence to support this benefit, and Major et al. (2002) found inconsistent results in their own study of Spanish and Chinese listeners.

Furthermore, there is a general consensus that unfamiliar accents, whether native or non-native, can negatively affect comprehension for all speakers (Flowerdew, 1994; Gass & Varonis, 1984; Major et al., 2002). However, exposure to and experience with a particular accent can aid our ability to recognise it and cope with its diverse realisations, ultimately meaning that the greater the exposure to a particular language variety, the easier comprehension becomes (Ballard & Winke, 2016; Smith & Bisazza, 1982). The role and influence of the media in establishing familiarity with diverse varieties, even when little or no face-to-face interaction takes place, was also suggested by Adank et al. (2009), who found that Scottish speakers based in Glasgow exhibited familiarity with and rapid processing of Standard British English, a variety ubiquitous across radio and television broadcasting throughout the UK. Yet, the reverse effect was not true for Standard British English speakers who were unfamiliar with the Glasgow accent, which is infrequently used in UK-wide media. They had slower response times and made more mistakes upon hearing Glasgow accent (Adank et al., 2009).

Attitude is another important factor to consider with regards to listener perception. As Holmes (2001) states, "people develop attitudes towards languages which reflect their views about those who speak the languages, and the contexts and functions with which they are associated" (p. 343). While it is well-documented that native speakers often negatively perceive deviations from the target language in L2 speaker speech, L2 speakers can also hold negative opinions of non-native speech (Abeywickrama, 2013; Dalton-Puffer et al., 1997; Kennedy & Trofimovich, 2010) and regional L1 speech (Archer, 2018). Such attitudes have been known to correlate with comprehension and perception of said accents—e.g., in ratings of how friendly or educated a speaker sounds—, and even influence comprehensibility (Eisenstein & Verdi, 1985). Indeed, Major et al. (2005) state unequivocally that the more prestigious a variety is, the greater the levels of comprehension will be.

3 The Study

The current study was conducted at the University of Strathclyde. The students who participated were expected to have been exposed to prestige models from prior English language instruction, from their ubiquity in pop culture and in the media, and from their own use of social media tools and converting platforms like TikTok, YouTube, and Instagram. This familiarity with prestige models could exert significant positive influence over these students' perception of prestige speaker accents. In contrast, it could afford negative characteristics to the speakers the students perceived to 'deviate' from more familiar prestige varieties, thus negatively affecting judgements of the L2 and regional L1 speakers. Based on the existing literature and the characteristics of this student population, three hypotheses were formulated to guide this investigation of L2 speakers' perceptions of various English accents:

1. Participants' familiarity and positive associations with prestige models may aid their identification, though the opposite is likely true with the SSE and L2 accents.
2. Participants will likely prefer prestige model accents (RP and GA) to SSE and L2 speakers' accents.
3. Participants will likely perceive and associate prestige models with more positive qualities than SSE or L2 models.

4 Methodology

4.1 Participants

Thirty students at the University of Strathclyde in Glasgow volunteered for this study. They were 20 males and 10 females between the ages of 20 and 38 years old. They were from Saudi Arabia (16), China (9), Libya (3), France (1), Kuwait (1), Thailand (1), and Colombia (1), and were enrolled on a mixture of programmes, including pre-entry English for Academic Purposes (EAP) courses, Masters, and Ph.D. studentships. As can be expected from the diverse pathways of study, their language level and time spent studying English varied significantly, from just three months to 28 years, with corresponding levels of English ranging from IELTS 4.5 to 7.5. The average length of residence in Glasgow was 4.4 months, but it ranged from new arrivals (one week) to two and a half years. Eleven participants had lived elsewhere in the UK prior to their move to Glasgow.

4.2 Test Procedure

After obtaining informed consent, a pre-test questionnaire gathered participants' background information, which included details such as their L1, home country, age, gender, last scores on a standardised English test (e.g., IELTS, TOEFL), length of English study, and length of residence in the UK and in Glasgow (if different). Next, participants were asked to identify who they communicated with more (L1 or L2 speakers of English) or if they talked a similar amount of time with both, and rate the ease with which they believed they could understand prestige models (RP/GA), and regional (SSE) and international (L2) accents using a 5-point Likert scale (1 = *Difficult or impossible to understand*; 5 = *Easy to understand*). A description was provided for each model to ensure students who were unfamiliar with the terms could understand the type of accent to which the questions referred. Then, participants completed a Verbal Guise test that measured their perceptual judgements of eight different speakers (see Sect. 4.3). Following the Verbal Guise test, participants were instructed to pick which of the eight accents they liked the most and which was the easiest to understand, and state why. Finally, participants were invited to a post-test interview to discuss the implications of the study within an English language teaching context. Five of the 30 participants completed the post-test interview.

4.3 Verbal Guise Test

Eight different speakers were recorded and used as the stimulus for participants' evaluations. The speakers were between 20 and 40 years old, educated, and all female in order to control for potential differences due to speaker gender. They were graduates, current research students, teachers, and other employees connected to English language teaching at the University of Strathclyde. Three speakers spoke SSE and were from Glasgow, Dundee, and Ayrshire. Two others had prestige model accents (RP and GA) and three were bilingual speakers with different L1 mother tongues (Spanish, Hindi, Mandarin). The speakers read the following four sentences aloud:

1. *The quick brown fox jumped over the lazy dog.*
2. *I just bought new shoes.*
3. *It's very warm in here.*
4. *I chose and bought the fruit carefully.*

Sentence one was an example taken from teaching materials. Sentences two to four were between 5 and 9 words long and were constructed around various lexical items containing phonological features which diverge in different accents, namely various vowel sounds, and /r/ in a post-vocalic position. The diversity of these features is particularly salient when comparing SSE and RP accents.

Participants were instructed to respond to the following questions using a 5-point Likert scale (1 = *I disagree 100%*; 5 = *I agree 100%*) after listening to each speaker:

Do you think the speaker.

- is a native speaker?
- has bad pronunciation?
- speaks too fast?
- has a foreign accent?
- is nice to listen to?
- has an annoying accent?
- would be a good English teacher?
- sounds educated?

How easy was the speaker to understand?

Test participants had no prior exposure to the test materials or speakers, and there were no limitations on the amount of times they could listen to each speaker. The test's design, requiring participants to scroll down to locate the next speaker, enabled distribution and distance between the different accents.

4.4 Data Analysis

To get an initial broad-spectrum view of the participants, their demographics, and perceptual judgements, descriptive statistics (including frequencies and means) were used. Following this, Chi-squared tests were undertaken to enable a comparison to be made between participants' accent preferences and their perceptions of ease of comprehension for each of the eight speakers. Pearson correlations were also used to establish if there was any association between the participants' ability to pinpoint speaker accents and any of the participants' variables. These variables included age, gender, level of proficiency, mother tongue, length of time studying English, length of time in the UK, and length of time in Glasgow. Finally, a Fischer exact test was used to corroborate results due to the small sample size ($N = 30$). Participants' anonymous responses are identified as P1, P2, P3, etc. A thematic analysis was conducted with the data from the post-test interviews. Pseudonyms are used to report participants' comments during the interviews.

5 Results

5.1 Pre-test Questionnaire Ratings

Participants' use of English with other speakers and their beliefs regarding which accents they found the most difficult/easier to understand were gathered prior to the Verbal Guise test. The majority of the participants (46%) said that they spoke to non-native English speakers/students more than native speakers, with 23.3% stating the opposite and 23.3% stating they spoke to both equally. Not unexpectedly, both RP and GA were generally rated as easy to understand, with over half of the respondents rating them either 'easy' or 'very easy' (RP = 50% and GA = 56.6%), and a large cohort awarding them a neutral mid-point rating (RP = 43.3%; GA = 30%). As per previous questions, justifications for these ratings were largely connected to clarity and familiarity with the models either through media or prior education. In contrast, ratings for SSE were more severe: 63.3% rated it as 'very difficult or impossible to understand' or 'difficult.' Some of the justifications for these ratings were as follows:

It is not very clear and sometimes fast. (P11)
It is very short and unclear. (P13)
It has a lot of strange pronunciations. (P10)
The accent is heavy. (P30)
It has a heavier sound. (P7)
We are not used to hear such accent. (P17)

Lastly, participants believed the bilingual or highly proficient L2 speaker option to be the easiest of all of the accents, with 60% rating it as 'easy' or 'very easy' and

26.6% as 'neutral.' A sense of camaraderie between some of the study participants and the 'L2 speakers' option became clear among their justifications for their ratings, with remarks made such as the following:

We are the same and use the same pronunciations. (P26)

We have the same process and we know what each other is talking about. (P21)

We learn and use the same words. (P19)

W*e are in the same level of pronunciation.* (P8)

However, other participants noted that that the intelligibility of an L2 speaker for them was still dependent on which country they came from and which accent they had.

These ratings at this stage of the study provided an important benchmark. This benchmark made it possible to draw comparisons between participants' beliefs regarding the different accents they were asked to listen to and their responses in the subsequent Verbal Guise test in which they were asked to identify the accent and then answer the questions.

5.2 Identification of Accent

Given so many students' strong positive feelings towards prestige models of English, it was thought prudent to establish if participants could correctly identify said models and those they perceived more negatively (i.e., SSE). The results showed that the most identifiable accent was the SSE accent from Glasgow, with 36.6% accurately pinpointing it. As all of the participants live in Glasgow, it was presumed that exposure to this model's phonological features could explain why it was more recognisable for some participants. However, no significant correlation was found between length of time living in Glasgow and ability to identify this accent ($p = 0.319$). The two other SSE speakers, neither of whom were from Glasgow, were not so easy for participants to identify, suggesting a particular salience in the Glasgow model's segmental or suprasegmental features, which participants seemed to recognise.

Indian was the next most identifiable accent (33.3% correct identifications), potentially due to its syllable-timed prosody, which stood out, particularly between the two native speaker accents that came before and after it during the Verbal Guise test. Following this was the GA accent (30% correct identifications), then RP (20% correct identifications). Interestingly, the majority of wrong answers for these two varieties were usually for each other, with 23.3% of participants labelling the RP accent as 'USA' and 16.6% believing the US accent to be from 'England' and 33.3% from 'the UK.' Such results suggest that while these varieties may be familiar, participants cannot always distinguish between the two. Among the remaining L2 accents, the Chinese accent was correctly identified by 20% of participants, all of whom were themselves Chinese native speakers. Using Pearson's correlations, a statistically significant score of $p = 0.004$ was obtained, thus determining that for the

Chinese accent, 'mother tongue' is a significant determiner for successful identification, in keeping with similar results found by Scales et al. (2006). Due to the demographics of the speakers and study participants, there was no other shared L1 between them to test further correlations of this feature.

The Spanish accent had little to no positive identification, perhaps because this speaker had had less discernible features associated with her mother tongue. Alternatively, it could be because participants had less interaction with speakers from Spain as this nationality is not the most prevalent among the University's international student population.

After examining variables such as length of time in the UK, length of time in Glasgow, length of time studying English, age, and gender, aside from shared mother tongue among Chinese speakers, only one other correlation was found regarding the ability to identify accents accurately: length of time in Glasgow and correct identification of the RP accent ($p = 0.027$). This may be due to the fact that many students begin intensive English instruction when they come to Glasgow, and this model remains prevalent in teaching resources, thus reinforcing exposure.

One significant issue which became visible from among the participants' results was an apparent lack of awareness of the geographical and phonological differences which exist in the individual nations within the UK: Scotland, Northern Ireland, Wales, and England. When asked to identify an accent, many participants simply wrote 'Great Britain' or 'UK,' suggesting a limited understanding of the very diverse accents in these nations, or a belief that there is only one 'British' accent in the whole of the UK. On the one hand, such beliefs are surprising, particularly given that many of the participants had been studying in Scotland, and in other parts of the UK, for months and even years, presumably absorbing some knowledge about their most recent country of residence. On the other hand, such results paralleled with a previous study conducted by the researcher (Archer, 2018) where among new arrivals to Glasgow from East and South East Asia, many admitted being unaware of any accent differences prior to coming to Scotland, assuming that everyone in Britain spoke 'British English'; many were disappointed to find this was not the case. As it stands, only answers referring to the individual nations such as 'Scotland' or 'England' were marked as correct and any use of 'UK' or 'Great Britain' as incorrect.

5.3 Accent Perceptions and Associated Qualities

To establish any qualities associated with each accent, participants were instructed to listen to each recording and then, against a list of statements (see Sect. 4.3), provide a rating between 1 and 5, where 1 = *I disagree 100%* and 5 = *I agree 100%* (see Table 1). The mean scores revealed interesting insights into the participants' perceptions of the eight accents. As predicted, RP elicited more generous feelings from participants, rating the highest among all the accents for 'nice to listen to,' 'would be a good English language teacher' and 'sounds educated.' In keeping with

Table 1 Mean accent ratings by speaker ($N = 30$)

Qualities	L2 Spanish	SSE Glasgow	GA	RP	L2 Indian	SSE Dundee	L2 Chinese	SSE Ayr
Bad pronunciation	2.5	1.9	1.8	1.9	3.0[a]	2.1	2.0	2.1
Speaks too fast	1.5	2.3	2.4	2.7	1.9	2.9[a]	2.2	2.1
Foreign accent	3.5	2.1	1.9	1.9	4.3[a]	2.2	2.7	2.7
Nice to listen to	3.5	3.5	3.9	4.2[a]	2.5	3.2	3.8	3.6
Annoying accent	2.1	1.9	1.9	1.9	2.7[a]	2.4	1.9	2.2
Would be a good EL teacher	3.1	3.4	4.1	4.3[a]	2.5	3.1	3.5	3.5
Sounds educated	3.4	3.3	3.7	3.9[a]	2.9	3.1	3.4	3.3
Easy to understand	4.0[a]	3.6	3.9	3.8	3.5	3.1	3.9	3.7

[a]Highest ratings

prior results, the GA speaker was also rated positively in these categories, similar to the RP speaker.

Among the L2 accents, the L2 Indian accent was rated highest for 'having a foreign accent,' 'having an annoying accent' and 'having bad pronunciation.' However, the rating for 'bad pronunciation' and 'annoying accent' were located midway on the scale, therefore interpretable as a more neutral or undecided response from participants regarding these particular qualities. The Chinese accent largely received neutral mid-point scores for most statements, except for 'has bad pronunciation' and 'has an annoying accent' with which participants seemed to disagree. Participants also felt this accent was 'easy to understand' and it received the same mean score as the prestige GA accent. The L2 Spanish accent was rated as the easiest to understand of all, and participants also clearly disagreed with the statement 'speaks too fast.' With regards to the SSE accents, in contrast with the pre-test views, participants' perceptions were significantly less severe. For the SSE Glasgow accent, participants tended to disagree with 'has bad pronunciation' and 'has an annoying accent,' which they rated identically to the RP accent. The remaining SSE accents' scores were also only slightly higher. The remaining ratings fell into the neutral mid-point, eliciting neither strong agreement nor disagreement.

5.4 Accent Preferences

The final task in the perception test instructed participants to choose which of the speakers' accents they liked the most, and which were the easiest to understand. They were allowed to listen again to each of the accent recordings to refresh their memories. As can be seen in Fig. 1, the accent that participants liked the most was the L2 Chinese accent, with eight participants (26.6%) selecting it and justifying their choice with comments such as "it was clear" and "it was familiar." As mentioned previously, many participants (40%) believed this to be a native accent, which may have also affected their perceptions of it. While examining the raw data, it became obvious that among those who selected the Chinese accent as their preference, there were no Chinese participants. In fact, the majority were from Saudi Arabia, along with one Colombian and one Thai. With regards to the Chinese participants, 10 out of 11 selected a prestige L1 model and one selected the Indian accent as their preference.

As predicted, the two prestige models were also popular among the listeners, with both being selected by seven participants each (23.3% each) and thus being the joint second favourite. Comments in favour of the RP accent ranged from "it's clear" and "it's easy" to "I think her voice is close to the British Standard" and "she speaks the best way." For those who selected the GA model, their justifications were unanimous in finding it the clearest and easiest to understand. Among the remaining accents, L2 Spanish was selected as the preferred model by three (10%), the SSE Glasgow model by two (6.6%), and the remaining L2s and L1 SSE models received 3.3%.

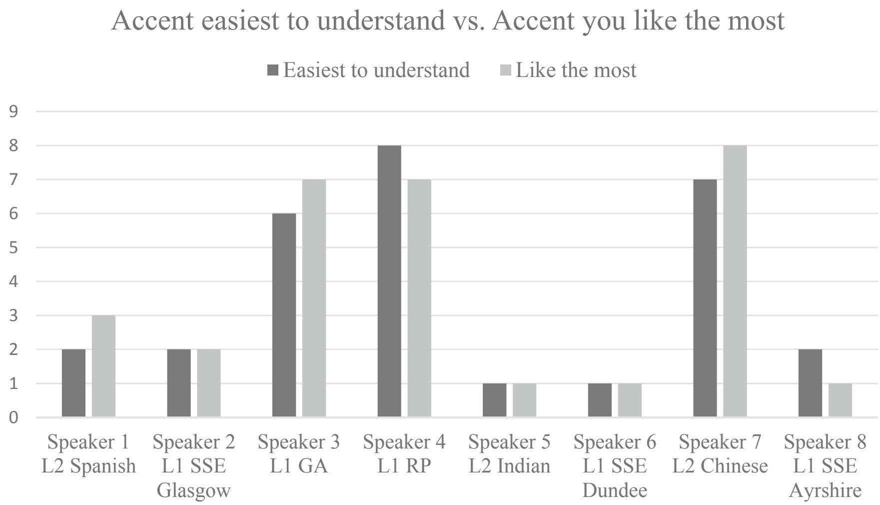

Fig. 1 'Easiest to understand' versus 'like the most' (preferred) accent

5.5 *Ease of Comprehension*

To ascertain if accent preference correlated with ease of comprehension, participants were also asked to select which accent they perceived to be the easiest to understand. The association between these two variables was then calculated using Phi. A statistically significant score of $p = 0.000$ was determined, thus establishing a clear relationship between the two variables, suggesting the easier an accent was perceived to be, the more it was liked (see Fig. 1). Further testing was undertaken to establish any possible correlation between mother tongue and the accent perceived as 'easiest to understand' (i.e., the 'Matched Interlanguage Speech Intelligibility Benefit'). In the current study, this correlation could only be checked among the Chinese participants, as this was the only nationality group present among both the speakers and listener participants. To establish if this benefit existed, a Pearson's chi-squared test was undertaken and it demonstrated that there was no significant correlation between participants' L1 and their perception of the easiest accent to understand ($p = 0.616$). Due to the small sample size in this study ($N = 30$), a Fisher's Exact Test was also used to corroborate the findings, confirming once again no correlation ($p = 0.604$).

6 Discussion

The purpose of this study was to investigate perceptions of L2 and L1 regional speech among international students, who are more likely to find prestige model accents (RP/GA) more familiar than other speech accents due to their prior education in their countries. Three hypotheses were formulated based on a review of literature. Each one is outlined and discussed in the following paragraphs.

Hypothesis 1 Participants' familiarity and positive associations with prestige models may aid their identification, though the opposite is likely true with the SSE and L2 accents.

This hypothesis was not confirmed. The results garnered show a limited number of participants being able to correctly identify the language background of the speaker, demonstrating that such a task is challenging for L2 learners. It also suggests that exposure alone may not be enough to make an accent recognisable; interaction with said accent may be necessary to enable learners' conceptualisation and subsequent recognition of its phonological features. Should that be the case, the identification of the SSE Glasgow and L2 Indian accents above all others could be due to prior interactional experiences the participants have had with them, as was found in the research of Austrian students by Dalton-Puffer et al. (1997).

Hypothesis 2 Participants will likely prefer prestige model accents (GA/RP) to SSE and L2 speakers' accents.

This hypothesis was not confirmed. Prestige model accents RP and GA have been reported as a preferred accent of L2 English students in numerous studies (Abeywickrama, 2013; Dalton-Puffer et al., 1997; Henderson et al., 2012). In contrast, in this study, the Chinese accent was selected as the most preferred, followed closely by both prestige models. An examination of the Chinese speaker's speech rate revealed that it was identical to that of the RP and was faster than the General American speakers. As faster speech rates are typically found in native and high proficiency L2 speakers (Munro, 1999), this feature could have contributed to the mistaken belief among participants that she was also a native speaker and thus was awarded the same positive associations. Aside from the fastest speaker (SSE Dundee), the next most preferred accents (RP, SSE Glasgow, GA) were also faster overall than all remaining models used. It could be therefore that accents believed to be native, potentially due to their speech rate, are more preferred.

Hypothesis 3 Participants will likely perceive and associate prestige models with more positive qualities than SSE or L2 models.

This hypothesis was confirmed. Mean scores of participant ratings revealed a tendency for prestige models RP and GA to be received more positively for sonority, perceived level of education, and potential to be a good English language teacher. Such results could simply be due to the fact that participants' familiarity with these models means less cognitive processing is required upon listening, resulting in faster and easier identification and comprehension.

7 Post-test Interviews

All participants were invited to a post-test interview to establish their beliefs about how international varieties might become more accepted and normalised among L2 students. Five participants volunteered to complete this final stage of the study. Four were Ph.D. students from Colombia, Thailand, Libya and China, and one was a Master's student from Taiwan.

Two themes emerged from a qualitative analysis of their responses: use of teaching materials, and approaches to accent in the classroom. Three of the participants believed using international voices in audio or video materials would be inspiring and normalising to L2 students, as shown in the following excerpts:

> *Show a video about a conference where people are speaking with different accents and sharing ideas in different accents and they can communicate without any problems. Sometimes, at least for me, I think that people won't understand me because of my accent, but when I have the chance to speak with other researchers from France or from Wales, they manage to understand me without any problems, because they are used to these different accents, especially in the academic world, it's very common.* (Hao, China)

> [Referring to a pronunciation class he had taken] *something that I really like is when we were watching interviews with people from different parts of the world. So for example, I remember this guy- the one who's really famous from South Africa, or the Malaysian guy*

that was interviewing the scientist in Florida, this kind of thing. So I think that would be a really good strategy, because at the end, there is not a right or wrong accent. It's not about this. It's a diverse world. It's totally globalised so we're having interaction with people from all parts of the world. So why people are just expecting to learn one English when they can just be conscious that there are various and the varieties extensive. That for me worked. (Jorge, Columbia)

I think every textbook only have one accent in common. Like in Taiwan, we have only American accent or British accent or Australian accent when we have a test or something. I think we have to be mixed because we could not choose the people we are talking about so we have to understand the different accent when we are learning. (Siyu, Taiwan)

Remarks such as these suggest that teaching materials and tasks could be useful ways of generating exposure to international Englishes and normalising accent diversity even prior to international university education.

Another of the interviewees highlighted the need to educate students about accent in a globalised world:

At school teach us there is no right accent. They have to change their mind. You have to communicate with, like, the global world, so you have to understand every accent. (Chalerm, Thailand)

Participants also remarked on the negative perceptions of diverse accents (namely the SSE accent) they encountered before arrival in Scotland, accessed online, or heard from peers.

My friend, she is studying in UK and she told me a lot like "oh you cannot understand the people in Scotland" but I think it's fine. (Siyu, Taiwan)

One participant felt strongly that negative perceptions could have significant implications for students considering further studies in Scotland.

All of student I think on YouTube or on Google for the accent in Scotland, all of them says "it's very difficult to understand; it's not good," and many students avoid coming to Glasgow to Scotland because of the reputation of the language, but when I came here I see the Scottish accent, especially in Glasgow is very nice, the Scottish accent has a rhythm like a melody when they speaking. (Amir, Libya)

Again, thoughtful pedagogical classroom tasks and teacher guidance could prevent, or at least, reduce the predictable negative reactions to unfamiliar sounds being generated by speakers with unfamiliar accents.

8 Implications

From the various stages of this research it can be seen that positive perceptions of prestige model accents, such as RP and GA, persist in English language learning despite the fact that L2 listeners may struggle to identify them as such. Moreover, said models are not essential for communication in the global environment where English is increasingly used as a lingua franca. As mentioned in the interviews, denying students opportunities to grow accustomed to diverse Englishes may limit their global vision, potentially even leading to life changing decisions, such as whether or not to visit, live or study abroad, despite the potential for learning and success they may find if they did. As Derwing et al. (2002) found, explicit instruction on the phonological features of Vietnamese accented English not only aided social work students' overall comprehension, but it also altered their perceptions of these speakers for the better. Therefore, providing input and support in class has the potential to alleviate some of the stigma attached to certain unfamiliar accents.

A simple way to begin the process of international English acclimatisation is via teaching resources. As gleaned from the existing literature and participants' comments during interviews, it seems clear that teaching resources play a considerable role in formulating what students believe is the acceptable or correct form of English. If major publishers continue to produce resources with limited or prestige models only, teachers could supplement their classroom practice with materials that represent the diversity of English as it is spoken worldwide. In an English for academic purposes context, this could be online lectures accessed from other universities, especially as many are now freely available. Alternatively, recordings of short concise academic presentations, such as those of the Three Minute Thesis competition, could be a simple way to provide diverse exposure, normalise accent variety, and even stimulate discussion on interesting topics and useful listening and decoding strategies.

9 Conclusion

The current study provides insight into the experiences of L2 English students studying in an international setting and contending with diverse novel L2 and regional L1 accents on a regular basis. Without the contextual clues provided in face-to-face communication, participants' observations of speech accents revealed that though many have developed a set of beliefs regarding prestige model native speaker accents, these beliefs do not necessarily help them identify or comprehend the language presented to them. Furthermore, without any information to guide them, some exert accent preferences contradicting their firmly held pre-listening beliefs. This can be seen in participants' preference for an L2 accent over a prestige model, or in participants' disagreement regarding SSE models having a 'bad pronunciation' (see Sect. 5.3), although the majority (63.3%) had previously stated that SSE was 'impossible' or 'difficult to understand' (see Sect. 5.1). Such results suggest that exclusive use of prestige models in ELT, and especially EAP, is unwarranted and could even be detrimental, affecting students' perceptions of accents, countries, and their inhabitants. However, an appropriate place in which acclimatisation to diverse Englishes can occur is in the classroom. With appropriate resources, teacher guidance can support students as they navigate their way through the diverse utterances and realisations of sounds, normalising the natural diversity that exists among international pronunciations of English.

References

Abeywickrama, P. (2013). Why not non-native varieties of English as listening comprehension test input? *RELC Journal, 44*(1), 59–74. https://doi.org/10.1177/0033688212473270

Adank, P., Evans, B. G., Stuart-Smith, J., & Scott, S. K. (2009). Comprehension of familiar and unfamiliar native accents under adverse listening conditions. *Journal of Experimental Psychology: Human Perception and Performance, 35*(2), 520–529. https://doi.org/10.1037/a0013552

Adank, P., & McQueen, J. M. (2007). The effect of an unfamiliar regional accent on spoken word comprehension. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 1925–1928). Pirrot. https://www.mpi.nl/publications/item59662/effect-unfamiliar-regional-accent-spoken-word-comprehension

Archer, G. (2018). *Pronunciation models in regional environments: A comparison and assessment of RP and SSE* (Unpublished MRes thesis). University of Glasgow. http://theses.gla.ac.uk/30983/

Baese-Berk, M. M., Bradlow, A. R., & Wright, B. A. (2013). Accent-independent adaptation to foreign accented speech. *The Journal of the Acoustical Society of America, 133*(3), EL174–EL180. https://doi.org/10.1121/1.4789864

Ballard, L., & Winke, P. (2016). 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. 259–271). Multilingual Matters. https://doi.org/10.21832/ISAACS6848

Bent, T., & Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. *The Journal of the Acoustic Society of America, 114*(3), 1600–1610. https://doi.org/10.1121/1.1603234

Buck, G. (2001). *Assessing listening*. Cambridge University Press. https://doi.org/10.1017/CBO9780511732959

Crowther, D., Trofimovich, P., & Isaacs, T. (2016). Linguistic dimensions of second language accent and comprehensibility: Non-native listeners perspectives. *Journal of Second Language Pronunciation, 2*(2), 160–182. https://doi.org/10.1075/jslp.2.2.02cro

Crystal, D. (2019). *The Cambridge encyclopaedia of the English language*. Cambridge University Press. https://doi.org/10.1017/9781108528931

Dalton-Puffer, C., Kaltenboeck, G., & Smit, U. (1997). Learner attitudes and pronunciation in Austria. *World Englishes, 16*(1), 115–128. https://doi.org/10.1111/1467-971X.00052

Derwing, T., Rossiter, M. J., & Munro, M. J. (2002). Teaching native speakers to listen to foreign-accented speech. *Journal of Multilingual and Multicultural Development, 23*(4), 245–259. https://doi.org/10.1080/01434630208666468

Eisenstein, M., & Verdi, G. (1985). The intelligibility of social dialects for working class adult learners of English. *Language Learning, 35*(2), 287–298. https://doi.org/10.1111/j.1467-1770.1985.tb01029.x

Flowerdew, J. (1994). Research of relevance to second language lecture comprehension-an overview. In J. Flowerdew (Ed.), *Academic listening* (pp. 7–29). Cambridge University Press. https://doi.org/10.1017/CBO9781139524612

Gass, S. M., & Varonis, E. M. (1984). The effect of familiarity on the comprehensibility of non-native speech. *Language Learning, 34*(1), 65–89. https://doi.org/10.1111/j.1467-1770.1984.tb00996.x

Goh, C. (1999). How much do learners know about the factors that influence their listening comprehension? *Hong Kong Journal of Applied Linguistics, 4*(1), 17–42.

Harding, L. (2008). Accent and academic listening assessment: A study of test-taker perceptions. *Melbourne Papers in Language Testing, 13*(1), 1–33. http://ltrc.unimelb.edu.au/mplt/papers/13_1_1_Harding.pdf

Henderson, A., Frost, D., Tergujeff, E., Kautzsch, A., Murphy, D., Kirkova-Naskova, A., Waniek-Klimczak, E., Levey, D., Cunningham, U., & Curnick, L. (2012). The English pronunciation in Europe survey: Selected results. *Research in Language, 10*(1), 5–27. https://hal.univ-smb.fr/hal-01644660

Holmes, J. (2001). *An introduction to sociolinguistics*. Pearson Education Limited. https://doi.org/10.4324/9780367821852

Jenkins, J. (2007). *English as a lingua franca: Attitude and identity*. Oxford University Press. https://global.oup.com/academic/product/english-as-a-lingua-franca-attitude-and-identity-9780194422376?lang=en&cc=fi

Kang, O., Vo, S. C. T., & Moran, M. K. (2016). Perceptual judgments of accented speech by listeners from different first language backgrounds. *TESL-EJ, 20*(1), 1–24. http://www.tesl-ej.org/wordpress/issues/volume20/ej77/ej77a1/

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

Kiczkowiak, M. (2021). Pronunciation in course books: English as a lingua franca perspective. *ELT Journal, 75*(1), 55–66. https://doi.org/10.1093/elt/ccaa068

Major, R. C., Fitzmaurice, F. B., & Balasubramanian, C. (2002). The effects of non-native accents on listening comprehension: Implications for ESL. *TESOL Quarterly, 36*(2), 173–190. https://doi.org/10.2307/3588329

Major, R. C., Fitszmaurice, F. B., Bunta, F., & Balasubramanian, C. (2005). Testing the effects of regional, ethnic, and international dialects of English on listening comprehension. *Language Learning, 55*(1), 37–69. https://doi.org/10.2307/3588329

Munro, M. (1999). The role of speaking rate in the perception of L2 speech. *The Journal of the Acoustical Society of America, 105*(2), 1032. https://doi.org/10.1111/1467-9922.00038

Munro, M., & Derwing, T. (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech, 38*(3), 289–306. https://doi.org/10.1177/002383099503800305

Munro, M., & Derwing, T. (1998). The effects of speaking rate on listener evaluations of native and foreign-accented speech. *Language Learning, 48*(2), 159–182. https://doi.org/10.1111/1467-9922.00038

Myhovych, I. (2019). International mobility as a means of insuring inclusive global higher education space. *Advanced Education, 12*, 80–86. https://doi.org/10.20535/2410-8286.137813

Perry, L. K., Mech, E. N., MacDonald, M. C., & Seidenberg, M. S. (2017). Influences of speech familiarity on immediate perception and final comprehension. *Psychonomic Bulletin & Review, 25*, 431–439. https://doi.org/10.3758/s13423-017-1297-5

Przedlacka, J. (2008). Models and myth updating the (non)standard accents. In K. Dziubalska-Kolaczyk & J. Przedlacka (Eds.), *English pronunciation models: A changing scene* (pp. 17–35). Peter Lang. https://www.phon.ucl.ac.uk/research/Przedlacka.pdf

Scales, J., Wennerstrom, A., Richard, D., & Wu, S. H. (2006). Language learners' perceptions of accent. *TESOL Quarterly, 40*(4), 715–738. https://doi.org/10.2307/40264305

Smith, L. E., & Bisazza, J. A. (1982). The comprehensibility of three varieties of English for college students in seven countries. *Language Learning, 32*(2), 259–269. https://doi.org/10.1111/j.1467-1770.1982.tb00971.x

Sung, C. C. M. (2016). Exposure to multiple accents of English in the English language teaching classroom: From second language learners' perspectives. *Innovation in Language Learning and Teaching, 10*(3), 190–205. https://doi.org/10.1080/17501229.2014.936869

Tauroza, S., & Luk, J. (1997). Accent and second language listening comprehension. *RELC Journal: A Journal of Language Teaching and Research in Southeast Asia, 28*, 54–71. https://doi.org/10.1177/003368829702800104

The British Council. (2013). *The English effect*. Retrieved February 17, 2019, from https://www.britishcouncil.org/sites/default/files/english-effect-report-v2.pdf

The University of Strathclyde. (n.d.). *Faculty of engineering*. Retrieved September 3, 2020, from https://www.strath.ac.uk/engineering/international/

Trofimovich, P., & Baker, W. (2006). Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition, 28*(1), 1–30. http://www.jstor.org/stable/44487037

Wells, J. C. (1982). *Accents of English*. Cambridge University Press. https://doi.org/10.1017/CBO9780511611766

Witteman, M., Weber, A., & McQueen, J. (2013). Foreign accent strength and listener familiarity with an accent codetermine speed of perceptual adaptation. *Attention, Perception, & Psychophysics, 75*, 537–556. https://doi.org/10.3758/s13414-012-0404-y

Zhao, Y. (1997). The effects of listeners' control of speech rate on second language comprehension. *Applied Linguistics, 18*(1), 49–68. https://doi.org/10.1093/APPLIN/18.1.49

Gemma Archer is Programme Coordinator and EAP teacher in the English Language Teaching Unit at the University of Strathclyde in Glasgow, Scotland. She is also Editor of *Speak Out!*—the journal of the IATEFL pronunciation SIG. Her research interests lie in the use of diverse forms of regional and global English in teaching materials and in the classroom.

Speech Assessment via Read-Alouds: A Critical Analysis of Diagnostic Passages

Takehiko Makino

Abstract My previous studies on Japanese speakers' pronunciation of English were based on recordings of TIMIT—a phonetic corpus that consists of isolated sentences with words chosen for the sake of phonological balance. The scope of possible analyses with TIMIT is limited because the recorded utterances are prosodically monotonous and the pronunciation of difficult words is unstable. To conduct more reliable speech assessments via read-alouds, I embarked on a critical analysis of diagnostic passages. This chapter critically reviews eight different diagnostic passages that have been used for data collection mainly in terms of phonetic coverage. My goal was to find one that meets the following criteria: contains every English phoneme; includes as many diphones as possible, especially those found to be difficult for Japanese speakers; does not contain words that are infrequent or too difficult for learners; constitutes a coherent passage which can elicit different prosodic patterns according to context; and illustrates a variety of speech acts that can elicit different intonation choices. Based on the critical review, the chapter argues for the selection of one diagnostic passage as the most suitable (after a few suggested adaptations) to collect English speech data from Japanese learners of English.

Keywords Learners' English pronunciation · Japanese speakers · Diagnostic passage · Phonetic coverage · Sentence types

1 Introduction

A few years ago, I developed *English Read by Japanese* (ERJ) *Phonetic Corpus* (Makino, 2013; Makino & Aoki, 2012) by using a small part of ERJ speech database (Minematsu et al., 2002)—a large collection (more than 70,000 files) of English sentences and words read aloud by 200 university students in Japan. Most of the ERJ sentences are based on TIMIT phonemically balanced set of sentences (Garofolo et al., 1993). The set of 800 files used for ERJ Phonetic Corpus was the same as

T. Makino (✉)
Faculty of Economics, Chuo University, Tokyo, Japan
e-mail: mackinaw@tamacc.chuo-u.ac.jp

V. G. Sardegna and A. Jarosz (eds.), *Theoretical and Practical Developments in English Speech Assessment, Research, and Training*, Second Language Learning and Teaching,
https://doi.org/10.1007/978-3-030-98218-8_6

that used in Minematsu et al. (2011), which studied the intelligibility of Japanese-accented pronunciation. I have used the ERJ Phonetic Corpus mainly to look at segmental patterns. However, while conducting my analyses, I found some significant drawbacks in the sentence set and the selected recordings in the Corpus. First, the phonemically balanced sentence set was originally devised for speech engineering. As a result, the set includes quite difficult words. According to the computation by Wordcounter.net (https://wordcounter.net/), the lexical items in the Corpus have a 11-12th grade reading level. This reading level can be very difficult for the average learner of English in Japan, and has probably led to unstable or erroneous pronunciations. Second, the sentence set consists of 420 isolated sentences (total 3,167 words) and many of them are short (8 words average). Because of this, most of the recorded utterances are prosodically monotonous, which has made it difficult to study different prosodic possibilities. Finally, the Corpus is not phonemically balanced for any particular speaker because only four out of the 120 sentences that comprise the balanced set for each speaker were randomly chosen. These limitations have led me to consider doing new recordings using a short passage.

The purpose of this chapter is to critically review different passages that have been used to collect learners' pronunciation of English so that I can find a more adequate data collection instrument for my study of Japanese speakers' English pronunciation. The chapter starts with an overview of the possible and attested problems in Japanese speakers' pronunciation of English. Then, it proposes the requirements for an ideal diagnostic passage. After these background discussions, it analyzes commonly used individual passages in terms of those requirements and identifies the most preferable choice. The chapter concludes with a discussion of the implications of this investigation.

2 Literature Review

2.1 Japanese Learners of English

The Japanese phonological system differs from that of English (Vance, 2008). It has a five-vowel system, and 14 phonemic consonants. Hence, its segmental inventory is less rich than that in English. It only has open syllables and virtually no initial consonant clusters. As for its prosody, lexical stress and sentence stress are always realized as falling pitch, but about half of the vocabulary does not have a lexical stress. Sentence intonation is realized as pitch movements at the end of intonation phrases (Venditti, 2005).

When pronouncing English vowels, Japanese speakers conflate /i/ and /ɪ/, /ɑː/ and /ɑɚ/ and /ɚː/, and /ʌ/ and /ə/, and several other pairs or sets. They also struggle with pronouncing English consonants /l, r, f, v, θ, ð/, and other consonants in certain phonetic contexts. For example, they tend to pronounce syllable-initial English /z/ as an affricate [dz], intervocalic English /dʒ/ as a fricative [ʒ], and intervocalic voiced plosives /b, d, g/ as fricatives [β, ð, ɣ]. They also do not distinguish word-final /z/ and /dz/. These are all cases of negative transfer from the Japanese pattern where voiced obstruents are usually realized as fricatives between vowels but as plosives or affricates word-initially. In my previous study on Japanese speakers' segmental patterns of English (Makino, 2013), voiceless plosives were also found to be spirantized in many cases, even though such patterns are not documented for spoken Japanese. I expect to find other "unexpected" patterns in new recordings.

Consonant clusters are also a major problem for Japanese speakers because they often insert a vowel between the consonants. *Strike* /straɪk/ becoming [sɯtoɾaikɯ] is one of their notorious mispronunciations, although in reality, the [ɯ] between voiceless consonants is usually dropped in spoken Japanese, so the more plausible mispronunciation is [stoɾaikɯ]. They also find difficult the pronunciation of word-to-word linking, especially between a final consonant and an initial vowel.

Japanese speakers are not good at placing English nuclear stresses on the appropriate syllables. They tend to use a rising pitch at the very end of yes–no questions even if the last syllable is unstressed and a continuous rise to the end from the earlier narrow focus is desirable (Ueyama, 1997), or place the nuclear stress on the last syllable even if it should be placed earlier in the intonation phrase. These errors may result from negative transfer from the already mentioned prosodic characteristics of Japanese. Surprisingly, such prosodic deviations have not been adequately documented in the literature even though they are sometimes discussed in informal observations. One of my motivations to collect Japanese speakers' reading of a passage in English is to be able to objectively describe the accented speech of Japanese learners of English.

2.2 Requirements for an Ideal Passage for a Read-Aloud Assessment

Ideally, a diagnostic passage should include the following:

- every phoneme of the target language, preferably in the same proportion as what occurs in authentic speech;
- as many diphones (types of two phoneme sequences) as possible, especially those found to be difficult for speakers of particular L1s (Japanese, in my case);
- a variety of sentence types and speech acts that elicit a range of prosodic and intonation patterns.

Additionally, the passage should *not*:

- contain words that are infrequent or too difficult for learners;
- be long—long passages may impose heavy burdens on the informants in recording sessions.

To my knowledge, only some of these requirements (most notably, the phoneme and diphone coverage) have been discussed in the construction of currently available passages (e.g., Hiki & Kakita, 2013; Kominek & Black, 2003). Although different sentence types may have been considered in the construction of the passages, I could not find a systematic review that discussed requirements for sentence types.

3 My Investigation

Before attempting to construct a passage that follows the ideal requirements just outlined, I set out to analyze the characteristics, advantages, and shortcomings of commonly used diagnostic passages in order to determine which could serve my data collection purposes most satisfactorily. This section first identifies the passages and criteria used for analysis, and then reports and discusses the findings.

3.1 Passage Selection

I have chosen the following passages for this survey:

- the "Stella" passage from the *Speech Accent Archive* (Weinberger, 2015);
- "The North Wind and the Sun" passage used for "Illustrations of the IPA" in the *Journal of the International Phonetic Association* (International Phonetic Association, 1999);
- "The Boy who Cried Wolf" passage, which was developed to improve upon "The North Wind and the Sun" (Deterding, 2006);
- the diagnostic passage in the *Manual of American English Pronunciation* (Prator & Robinett, 1984);
- the diagnostic passage in *Teaching Pronunciation* (Celce-Murcia et al., 2010);
- the diagnostic passage in *Well Said* (Grant, 2017)—a widely used textbook;
- the short version of "Arthur the Rat," reproduced in *A Course in Phonetics* (Ladefoged & Johnson, 2015);
- "Text for phonemic contrasts" in William Labov's study of New York speech (Labov, 2006, originally in 1966).

The passages I have chosen are by no means all the ones that have been used to collect learners' pronunciation of English. There are at least three other important sentence sets that I could have included: "The Rainbow passage" (formerly used) and the "Comma Gets a Cure" (currently used) by the online International Dialects of English Archive (Maier, 2019), and the Arctic sentence set (Kominek & Black, 2003).

"The Rainbow passage" consists of 331 words of 9th-10th grade level, with an average sentence length of 18 words. "Comma Gets a Cure" has 372 words in 9th-10th grade level, with an average sentence length of 17 words. Critics have found these two passages rather "unnatural," perhaps because words were chosen to represent all the English phonemes in terms of "standard lexical sets" (Wells, 1982) that could cover all the possible (vowel) contrasts in different standard native-speaker varieties. Because unnatural passages can produce unnatural utterances, especially as regards prosody, I have chosen not to analyze these two passages. Another major reason is their lack of interrogative sentences, which are necessary for collecting different uses of question intonations.

The Arctic sentence set was originally produced for use in the development of speech synthesis. Importantly for us, it has been used for the collection of L2-Arctic Corpus (Zhao et al., 2018), a collection of L2 English speech by (currently) 24 speakers whose L1s are Arabic, Mandarin Chinese, Hindi, Korean, Spanish, and Vietnamese. This set consists of 1,132 "sentence prompts" (whose average length is 9 words) with 9,998 words (in 9th-10th grade level). One of its major merits is the diphone coverage at 79.6%. Although I have not computed the diphone coverages for the passages assessed in this chapter, they must be much smaller because the number of possible diphones in English is 1,610, according to the developers of the Arctic set, and the numbers of patterns I present below in the analysis are in lower hundreds at most. However, the set's major merit is also one of its drawbacks. This large set is probably too burdensome for speakers to comfortably read aloud in the recording session. Another problem derives from the way the developers chose sentences for their prompts. Although the set draws on a running literary text, they "pruned" it automatically and manually so that the prompts meet their requirements about length, pronounceability, and types of words and grammar. As a result, the set is largely a collection of isolated sentences rather than a coherent text. As I argue in this chapter, this is not desirable for collecting uses of different prosodic patterns according to contexts.

3.2 Data Analysis

The analysis of each passage starts with a general description in terms of its author, number of words, reading level according to Wordcounter.net (https://wordcounter.net/), and average sentence length. What follows is a descriptive linguistic analysis of the range of phonemes (based on the General American phonemic inventory), consonant clusters, word-to-word sound combinations (excluding vowel-to-consonant combinations[1]), and sentence types embedded in each text. The identified sounds and sentence types constitute potential candidates for speech analyses of read-aloud speech. The findings from the linguistic analysis are then critically discussed in view of the advantages and shortcomings they pose for speech data collection and analysis.

[1] Vowel-to-consonant combinations are not considered because they are not candidates for connected speech phenomena, such as linking and assimilation, and therefore do not generally pose pronunciation problems for language learners.

3.3 Results and Discussion

3.3.1 "Stella"

The passage "Stella" has been used to collect speech at the website *Speech Accent Archive* (Weinberger, 2015). The collection includes recordings from speakers of as many as 386 different first languages. The text is only 69 words long (with 55 unique word forms). According to Wordcounter.net, the text has a 5th-6th grade reading level. The average sentence length is 18 words. The linguistic analysis revealed that this passage included the following candidates for speech analysis:

- All phonemes except /ʒ, dʒ, j, aʊ, ɪɚ, ɛɚ, ɑɚ, ʊɚ/.
- 23 types of word-internal consonant clusters ("#" denoting word boundaries):
  - initial (12): #bl, #br, #fr, #pl, #sk, #sl, #sm, #sn, #sp, #st, #tr, #θr;
  - final (8): bz#, ts#, gz#, ks#, nd#, nz#, ŋz#, sk#;
  - intervocalic medial (3): ls, nt̬, nzd.
- 33 types of word-to-word sound combinations:
  - vowel to vowel (3): i#ɔ, i#ə, aɪ#ə;
  - consonant to vowel (3): k#ə, z#ɪ, z#ə;
  - possible t/d-flapping (1): d#ə;
  - possible place assimilation (2): d#b, d#m;
  - other consonant to consonant (24): p#ð, t#ð, t#h, d#w, k#f, k#s, k#h, g#t, g#f, v#b, v#f, v#θ, ð#h, s#s, z#k, z#θ, z#w, ʃ#s, m#ð, n#s, ŋ#ð, l#p, l#g, l#s.
- 4 sentence types:
  - statements (with falling tones): We also need a small plastic ↘snake…;
  - commands (with falling tones): Please call ↘Stella;
  - non-final intonation phrase (rises): She can scoop these things into three red ↘bags↗…;
  - a list (with non-final rises and a final fall): Six spoons of fresh ↗snow peas ‖ five thick slabs of ↗blue cheese ‖ and maybe a ↘snack.

The short length of this passage makes it handy for collecting samples from a large number of people from different language backgrounds, and that is the main purpose of the *Speech Accent Archive*. However, this passage is far from adequate for collecting speech data for a phonetic study because most of the requirements identified in Sect. 2.2 for an ideal diagnostic text are not fulfilled. Most importantly, some phonemes are not included, and the coverage of consonant clusters and word-to-word combinations is very limited. Also, the majority of its sentences is commands and statements, which will only elicit falling intonation. Although the list included in the text will probably elicit rising tones, the major use of rises in questions is not represented. Hence, the reading will be monotonous, to say the least, and the passage will not be useful for assessing prosody.

3.3.2 "The North Wind and the Sun"

The passage "The North Wind and the Sun" was first used to demonstrate the uses of IPA symbols in "Illustrations of the IPA," a major section in the *Journal of the International Phonetic Association.* The text has been translated into different languages to demonstrate the uses of the alphabet in different languages. As such, it was not designed for collecting speech samples with adequate phonetic balance. The English version consists of 113 words (66 unique word forms) in the 9th-10th grade level. The average sentence length is 23 words. The linguistic analysis revealed the following regarding representative features:

- All phonemes except /ʒ, ɑ, ɔɪ, ɪɚ, ɛɚ, ʊɚ/.
- 20 types of word-internal consonant clusters:
  - initial (4): #bl, #kl, #str, #tr;
  - final (6): dʒd#, ld#, mpt#, nd#, pt#, st#;
  - intervocalic medial (10): bl, gr, ks, ml, nf, ns, ŋg, sl, sp, tl.
- 45 types of word-to-word sound combinations:
  - vowel to vowel (4): i#ʌ, i#ə, u#ə, eɪ#ə;
  - consonant to vowel (7): k#ɔ, k#ə, t#ɪ, d#ə, z#ə, v#ʌ, m#ə;
  - n- and r-linking (3): n#aʊ, n#ə, ɚ#ə;
  - possible t/d-flapping (1): d̬#ɪ;
  - lateral release (1): t#l;
  - possible place assimilation (2): d#b, d#g;
  - other consonant to consonant (27): p#ð, t#ð, t#s, t#w, d#ð, d#h, d#s, d#t, d#w, k#h, f#h, f#ʃ, v#ð, θ#w, s#ð, z#ð, z#h, z#k, tʃ#w, m#k, n#ð, n#ʃ, n#h, n#w, ŋ#ð, ŋ#r, ŋ#w.
- 2 sentence types:
  - statements (falls): the North Wind gave up the at↘tempt;
  - non-final intonation phrases (rises): ↘Then↗ ‖ the North Wind…

I know of at least one L2 speech corpus project (AESOP corpus of Asian Englishes by Meng et al., 2009) which makes use of this passage for collecting speech samples, but the nature of the text (especially the lack of some phonemes and monotonous intonation) can limit its usefulness for speech data analysis. Importantly, the text is far from ideal for obtaining a variety of prosodic patterns given that it only includes statements, which will mostly elicit falling intonation patterns apart from possible rises in sentence-medial phrasings.

3.3.3 "The Boy Who Cried Wolf"

"The Boy who Cried Wolf" (Deterding, 2006) was developed to eliminate deficiencies of "The North Wind and the Sun" passage by collecting phonetically balanced speech samples with fewer repetition of the same words (Hiki & Kakita, 2013). It consists of 216 words (134 unique word forms) in the 7th-8th grade level, and the average sentence length is 27 words. The linguistic analysis revealed the following regarding representative features:

- All phonemes (despite being originally devised for standard British pronunciation).
- 36 types of word-internal consonant clusters:
  - initial (7): #fl, #fr, #kr, #pl, #st, #tr, #θr;
  - final (17): kst#, ldz#, lf#, mz#, nd#, ndʒ#, ns#, nst#, nt#, ntn#, pt#, st#, ʃt#, t̬l#, tn#, vn#, znz#;
  - intervocalic medial (12): ft, gz, ks, ktl, ktʃ, ls, mp, ms, ns, sf, st, tl.
- 97 types of word-to-word sound combinations:
  - vowel to vowel (6): i#ɛɚ, eɪ#ə, ɔɪ#ə, oʊ#oʊ, oʊ#ə, u#ə;
  - consonant to vowel (15): p#ə, t#ɪ, d#ɔ, k#ə, v#ə, s#ɪ, s#ə, z#i, z#ɔ, m#ɪ, m#aʊ, m#ə, ŋ#æ, ŋ#ɪ, l#ə;
  - possible t/d-flapping (7): t̬#i, t̬#æ, t̬#ə, t̬#ʌ, d̬#ə, d̬#ɪ, d̬#aʊ;
  - n- and r-linking (3): n#ə, ɪɚ#ə, ɚ#ə;
  - lateral release (2): t#l, n#l;
  - possible place assimilation (5): t#b, d#p, d#b, d#k, n#b;
  - other consonant-to-consonant combinations (59): t#f, t#h, t#n, t#s, t#t, d#t, d#d, d#dʒ, d#h, d#s, d#w, k#f, f#ð, f#h, f#k, f#w, v#b, v#ð, v#h, v#k, v#tʃ, ð#h, s#g, s#k, s#m, z#ð, z#f, z#k, z#l, z#s, z#t, z#v, z#w, ts#f, ts#j, tʃ#h, tʃ#p, dʒ#f, dʒ#ʃ, m#t, m#k, m#f, m#ð, n#t, n#d, n#f, n#ð, n#s, n#h, ŋ#t, ŋ#d, ŋ#f, ŋ#ʃ, ŋ#h, ŋ#w, l#d, l#f, l#ð, l#r.
- 4 sentence types:
  - statements (falls): the wolf had a ↘feast;
  - commands (falls): Go a↘way;
  - non-final intonation phrases (rises): As soon as they ↘heard ↗him;
  - calls (falls): ↘Wolf || ↘wolf.

This passage was constructed for use in phonetic studies. Hence, it has some clear advantages for phonetic analyses: all the phonemes are represented, and there are more word-to-word combinations for its relatively short text. Also, the sentences in the text are relatively long (the longest among the passages discussed in this chapter), so they are more likely to be divided into intonation phrases, which can carry non-final (rising) tones. However, the passage does not seem to have been designed with prosody in mind as it does not contain questions, which would elicit sentence-final rising intonation. This is a major omission. Although the passage has reasonably animated content which could elicit more expressive performance (i.e., wider pitch ranges), it is clearly more adequate for segmental studies than for collecting data on different prosodic patterns.

3.3.4 *Manual of American English Pronunciation* Diagnostic Passage

The Manual of American English Pronunciation (Prator & Robinett, 1984) is a textbook which was widely used during the latter half of the twentieth century in programs where the target accent was American. It contains a diagnostic text that can be used by the instructor to assess English learners' pronunciation. The passage has 165 words in the 9th-10th grade level, and the average sentence length is 15 words. The linguistic analysis revealed the following regarding representative features:

- All phonemes except /ɪɚ, ɑɚ, ɔɪ/.
- 32 types of word-internal consonant clusters:
  - initial (7): #pr, #tr, #dr, #fr, #kw, #sp, #st;
  - final (10): dz#, dn̩t#, ks#, kt#, mz#, nt#, nd#, ŋk#, nz#, lf#;
  - intervocalic medial (15): pr, bl, mp, nt̬, ntr, ms, nf, ns, ŋgw, st, stʃ, dv, kt, ltʃr, dn̩l.
- 45 types of word-to-word sound combinations:
  - vowel to vowel (1): i#ɪ;
  - consonant to vowel (10): t#ɪ, d#ɔ, d#aʊ, k#æ, k#ə, z#ɪ, dʒ#ə, l#æ, l#ə, l#ɪ;
  - possible t/d-flapping (4): t̬#i, t̬#ɪ, t̬#ɚ, d̬#ɪ;
  - lateral release (1): t#l;
  - r-linking (4): ɚ#ɔ, ɚ#ə, ɚ#ɪ, ɔɚ#ɪ;
  - other consonant-to-consonant combinations (25): p#s, t#d, t#f, t#s, t#h, t#r, d#f, d#s, d#h; f#k, f#ð, v#h, s#t, s#l, z#t, z#ð, z#m, z#w, m#t, m#dʒ, ŋ#d, l#p, l#b, l#s, l#ʃ.
- 8 sentence types:
  - statements (falls): All of this will take will ↘power;
  - yes-no questions (rises): Should he spend all of his time ↗studying?;
  - wh-questions (falls): Where should he ↘live?;
  - alternative questions (a medial rise and a terminal fall): Would it be better if he looked for a private room ↗off campus, || or if he stayed in a ↘dormitory?;
  - a yes–no question spanning some intonation phrases (a sequence of rises): Shouldn't he try to take ad↗vantage || of the many social ac↗tivities || which are ↗offered?);
  - non-final intonation phrases (rises): When a student from another country comes to study in the United ↗States…;
  - tag questions (falls or rises): …doesn't develop ↘suddenly, || ↘does it?;
  - vocatives (rises): But let me ↘tell you, || my ↗friend.

This passage contains a wide range of sentence types. However, in some cases, the plausible intonation patterns are rather unusual and difficult. For example, the sequence of rises in the question spanning three intonation phrases is hard to pronounce for some, and probably not common enough. One may want to collect simpler patterns before delving into that sort of prosody. Another possible objection is that the sentences sound dated. For instance, using "he" for a person unspecified for gender does not feel correct anymore. Because the omission of some phonemes is also a problem, the text needs to be revised before it can be chosen for data collection.

3.3.5 *Teaching Pronunciation* Diagnostic Passage

Teaching Pronunciation (Celce-Murcia et al., 2010) is one of the major pronunciation textbooks aiming at prospective teachers. Its diagnostic passage has 226 words in the 9th-10th grade level, and the average sentence length is 12 words. The linguistic analysis revealed the following regarding representative features:

- All phonemes except /ɔɪ/.
- 42 types of word-internal consonant clusters:
  - initial (6): #pr, #fr, #θr, #str, #kw, #sp;
  - final (13): kt#, st#; ns#, nt#, nts#, ntʃ#; nd#, ndz#, ndʒd#; ŋz#; lt#, lz#; pl̩#;
  - intervocalic medial (23): ks, ksp, kt; gz, gn, gr; stʃ, sn; mp, mpr, mb; ntr, nd, nt̬, nf, nfl, ns, nl; ŋgl, ŋgw; ls, ld, ldr.
- 75 types of word-to-word sound combinations:
  - vowel to vowel (4): i#ɔ, oʊ#aɪ, u#ɔ, aɪ#ɪ;
  - consonant to vowel (17): d#ɪ, t#ɪ, t#oʊ, t#ə, v#ʌ, s#ɔɚ, z#ɪ, z#ɔ, z#oʊ, ʃ#ɚ, ts#ə, tʃ#æ, dʒ#ɪ, m#ə, ŋ#æ, ŋ#ə, l̩#ə;
  - n- and r-linking: (6): n#æ, n#ə, n#ɪ, ɚ#ə, ɚ#ɪ, eɚ#ɚ;
  - possible t-flapping (2): t̬#ə, t̬#ɪ;
  - possible coalescence (1): t#j;
  - possible place assimilation (3): t#p, t#b, d#p;
  - lateral release (1): d#l;
  - other consonant-to-consonant combinations (41): t#t, t#d, t#f, t#tʃ, t#m; d#t, d#d, d#h, d#w; k#p, k#s, k#h, k#w; f#n; v#l, v#s, ð#p; s#b, s#k; z#p, z#t, z#k, z#f, z#ð, z#j; ʃ#s, ʃ#j; dʒ#w, dʒ#j; n#t, n#k, n#ð, n#l, n#w; ŋ#k, ŋ#t; l#t, l#j; l̩#f, l̩#ð, l̩#r.
- 9 sentence types:
  - statements (falls): There are a couple of ↘answers to this question;
  - yes-no questions (rises): Is English your native ↗language?;
  - wh-questions (falls): Why is it difficult to speak a foreign ↘language without an accent?;
  - non-final intonation phrases (rises): If ↘not↗…;

– alternative questions (a medial rise and a final fall): Will you make ↗progress, || or will you give ↘up?;
– final comment clauses (low levels): Only ↘time will tell, || I'm a → fraid;
– parentheticals (rises): for e↘xample↗;
– lists (non-final rises and a final fall): concentrated hard ↗work || a good ↗ear || and a strong ↘ambition;
– strong assertions (a wider pitch range): You can im↗↘prove!

This passage contains a wide range of sentence types. The sentences are shorter and less complex than those by Prator and Robinett (1984), and this is clearly an advantage in eliciting straightforward intonation patterns that the speaker has acquired. The incomplete phonemic coverage could be a problem, but the omission of only /ɔɪ/ may not be a major concern because it is not one of the difficult vowels of English, and a word containing /ɔɪ/ could easily be added to the text.

3.3.6 *Well Said* Diagnostic Passage

Well Said (Grant, 2017) is probably one of the most widely used American pronunciation textbooks currently in print. The diagnostic passage has 138 words (89 unique word forms) in the 11th-12th grade level, and the average sentence length is 12 words. The linguistic analysis revealed the following regarding representative features:

- All phonemes except /ʒ, ɔɪ, ɛɚ, ɑɚ, ʊɚ/.
- 32 types of word-internal consonant clusters:
  – initial (4): #kl, #pl, #pr, #sp;
  – final (14): ks#, kt#, lt#, lts#, mz#, nd#, ndz#, nt#, nz#, ŋz#, tnt#, tʃt#, zn#, znz#;
  – intervocalic medial (14): gr, ks, ksp, ldr, ltʃr, mb, mp, mpl, mpr, ns, nstr, ŋgl, ŋgw, sp.
- 66 types of word-to-word sound combinations:
  – vowel to vowel (6): i#ɪ, u#ɚ, u#ɛ, eɪ#ɚ, eɪ#ɪ, aɪ#ɪ;
  – consonant to vowel (15): t#ɪ, d#ɪ, k#ə, v#ə, s#ɪ, z#ɪ, z#ɔ, z#ə, z#ɚ, dz#ə, dʒ#ɪ, m#ɛ, ŋ#ɪ, l#ɪ, l#ɚ;
  – n- and r-linking (5): n#ə, n#ɪ, ɔɚ#ɛ, ɚ#æ, ɚ#ə;
  – possible t/d-flapping (5) t̬#ɪ, t̬#ɛ, t̬#ɔ, d̬#ɪ, d̬#ə;
  – nasal release (1): t#n;
  – lateral release (2): t#l, d#l;
  – possible coalescence (1): z#j;
  – other consonant-to-consonant combinations (31): t#t, t#k, t#f, t#ð, t#j, t#r, d#k, d#s, d#h, f#t, f#j, v#g, v#ð, v#s, v#j, θ#m, ð#ð, s#l, s#m, z#p, z#s, z#h, z#n, z#w, ts#s, dʒ#ð, dʒ#s, n#w, ŋ#t, ŋ#s, ŋ#tʃ.
- 5 sentence types:
  – statements (falls): pronunciation of a new language is not auto↘matic;

- yes-no questions (rises): Have you ever watched young ↗children ‖ play with the sounds of the languages they are ↗learning;
- wh-questions (falls): Why is progress in adults more ↘limited?;
- non-final intonation phrases (rises): For young ↘child↗ren;
- lists (rises): They ↘imi↗tate, ‖ re↘peat↗, ‖ and sing sound combi↘nations without effort.

The phonemic coverage is incomplete in this passage. Supplying the missing five phonemes may not be an easy task. One might think that the vowels /ɛɚ, ɑɚ, ʊɚ/ are combinations of /ɛ, ɑ, ʊ/ plus /r/ phonologically and, hence, not absolutely necessary for pronunciation assessment. However, Japanese speakers perceive postvocalic r's as vowels which sound quite different from prevocalic r's, so it is necessary to collect how Japanese learners of English pronounce them in these combinations.

The coverage of sentence types is fair but not good enough, either. Having yes–no and wh- questions barely fulfills the minimum requirement. The lack of alternative questions is a major omission, and I would like to collect more types of speech acts such as those found in the passage from Celce-Murcia et al. (2010) (see Sect. 3.3.5). This passage is probably too short to include other types of speech acts.

3.3.7 "Arthur the Rat" Short Version

The original version of "Arthur the Rat" (consisting of 594 words) was devised by Henry Sweet and used extensively in the fieldwork for the *Dictionary of American Regional English* (Cassidy, 1985). The shortened version of the original, which is reproduced in *A Course in Phonetics* (Ladefoged & Johnson, 2015), has 339 words (197 unique word forms). Its vocabulary is in the 5th-6th grade level, and the average sentence length is 22 words. The linguistic analysis revealed the following regarding representative features:

- All phonemes except /ʒ, ʊɚ/, which are the most infrequent consonant and vowel in English.
- 40 types of consonant clusters:
  - initial (8): #bl, #fl, #fr, #gr, #kr, #sk, #st, #tr;
  - final (21): dn#, dnt#, ft#, kt#, ld#, ld#, lf#, lm#, lz#, md#, nd#, ns#, nt#, ŋk#, skt#, st#, sts#, t̬l#, tn#, tʃt#, vd#;
  - intervocalic medial (11): ft, lw, ml, ms, nd, ndl, nl, ns, nt̬, ŋgr, sl.
- 119 types of word-to-word sound combinations:
  - vowel to vowel (10): oʊ#aʊ, i#æ, oʊ#i, u#ɪ, i#ɑ, eɪ#ə, ɔ#ɑ, u#ɚ, i#oʊ, i#ə;
  - consonant to vowel (23): p#ə, t#ə, d#ɔ, d#ɑɚ, d#aʊ, d#ə, d#ð, k#ʌ, k#ə, k#aɪ, f#aʊ, f#ɪ, v#aʊ, ð#ə, s#ɔ, s#ə, s#ɚ, z#ɪ, z#æ, z#ə, dz#æ, m#ə, ŋ#ə;
  - possible t/d-flapping (5): t̬#ə, d̬#ɔ, d̬#ʌ, d̬#oʊ, d̬#ə;
  - n- and r-linking (11): n#aɪ, n#ə, n#ɛ, n#ɪ, n#j, ɔɚ#ɪ, ɪɚ#ə, ɚ#ɒ, ɚ#ə, ɚ#ɪ, ɚ#ʌ;
  - nasal release (2): t#n, d#n;

  - lateral release (3): t#l, d#l, n#l;
  - possible assimilation (4): t#k, d#b, d#k, d#g;
  - possible coalescence (1): n#ð;
  - other consonant-to-consonant combinations (60): p#h, t#d, t#ð, t#f, t#s, t#h, t#w, d#t, d#f, d#h, d#w, k#t, k#ð, k#h, k#m, v#g, v#ð, v#n, v#w, ð#f, ð#s, ð#h, d#r, s#ð, s#m, s#n, s#w, z#d, z#g, z#f, z#ʃ, z#h, z#m, z#r, z#w, ts#k, ts#g, ts#s, ts#h, tʃ#f, m#b, m#t, m#d, m#h, m#m, m#n, n#f, n#s, n#h, n#r, n#w, ŋ#t, ŋ#l, ŋ#r, l#t, l#g, l#ð, l#s, l#h, l#r.

- 5 sentence types:

  - statements (rises): there was a young rat ↘Arthur…;
  - commands (falls): Now look ↘here;
  - tag questions (falls or rises): You're ↘coming, || of ↗course?;
  - non-final intonation phrases (rises): One rainy ↘day↗;
  - calls (falls): Right about ↘face.

This passage lacks the most infrequent phonemes /ʒ, ʊɚ/. In its original longer version, the only missing phoneme is /ʊɚ/. It is surprising that a passage extensively used for dialect fieldwork does not include all phonemes. However, it is possible to add words like *usual(ly)* and *poor*, *pure* or *cure* to cover all phonemes. The coverage of consonant clusters and word-to-word combinations is the best of all the texts reviewed in this chapter probably because "Arthur the Rat" is longer than most of them, except the one by Labov (2006) analyzed in the next section. Perhaps the biggest weakness of this passage is the coverage of sentence types: wh-questions and alternative questions are missing. These omissions are problematic because it is impossible to include other sentence types without revising the text extensively. Finally, the story itself does not seem to be exciting enough.

3.3.8 Labov's "text for Phonemic Contrasts"

Bill Labov used this passage for his study of New York speech (Labov, 2006) in the 1960s, but I do not know of any other study utilizing it. It consists of 23 sentences with 349 words (212 unique word forms) in the 9th-10th grade level. The average sentence length is 15 words. The linguistic analysis revealed the following regarding representative features:

- All phonemes.
- 44 types of word-internal consonant clusters:

  initial (10): #br, #fr, #gr, #pl, #sl, #sm, #st, #str, #sw, #θr;
  final (18): bl#, dʒd#, fθ#, kst#, ld#, lf#, lk#, lm#, lz#, nd#, nt#, nz#, sk#, skt#, st#, t̬l#, zd#, znt#;
  intervocalic medial (16): bm, bw, dtw, ktʃ, mw, nd, ndr, nf, nh, nt̬, ŋg, sk, st, tnl, tns, tʃr.

- 110 types of word-to-word sound combinations:

- vowel to vowel (6): oʊ#ə, i#ə, aɪ#ɑ, i#ɛ, i#æ, u#æ;
- consonant to vowel (20): p#ə, t#ɒ, t#ɪ, t̬#ə, f#ə, v#ə, θ#ɪ, s#aɪ, s#ʌ, s#ə, s#ɚ, z#ə, ʒ#ɑ, m#ʌ, ŋ#ɑ, ŋ#ə, ŋ#ə, l#ɪ, l#oʊ, l#ə;
- n-linking (4): n#aɪ, n#ə, n#eɪ, n#ɛ, n#j;
- r-linking (2): ɚ#ɪ, ɛɚ#æ;
- nasal release (2): p#m, d#n;
- lateral release (3): t#l, d#l, n#l;
- place assimilation possible (11): n#b, n#d, n#k, n#m, t#p, t#k, t#ɡ, t#m, d#b, d#ɡ, d#m;
- coalescence possible (1): n#ð;
- other consonant-to-consonant patterns (63): p#ʃ, p#θ, t#d, t#ð, t#f, t#h, t#s, t#ʃ, t#t, t#tʃ, t#w, t#θ, d#ð, d#h, d#s, d#w, k#ð, k#f, k#k, k#l, k#m, k#s, k#t, v#b, v#ð, v#j, v#v, θ#m, θ#s, ð#k, ð#t, s#f, s#h, s#m, z#m, z#s, z#ʃ, z#t, z#θ, tθ#æ, ts#b, ts#ð, ts#dʒ, ts#t, ts#w, tʃ#w, dʒ#ʃ, m#ð, m#f, m#k, m#t, m#w, n#n, n#r, n#s, n#ʃ, ŋ#f, ŋ#s, ŋ#w, l#ð, l#l, l#r, l#w.

- 10 sentence types:
  - statements (falls): Mary got her ↘finger in the pie;
  - yes-no questions (rises): Are they running ↗submarines to the Jersey shore?;
  - wh-questions (falls): And what's the source of ↘your information;
  - commands (falls): Don't tell this man any fairy ↘tales about a ferry;
  - non-final intonation phrases (rises): When Mary starts to sound ↘humor↗ous,…;
  - vocatives (rises): "And what's the source of your infor↘mation, ↗Joseph?";
  - calls (wide falls): My ↘God!;
  - strong assertion (wide falls): Oh yes he ↘can!;
  - reported speech: "You're certainly in the ↘dark," I ↘told her;
  - irony: They tore down that dock ten ↘years ago, when you were in ↘diapers.

This passage has a complete phoneme coverage, the largest number of consonant clusters and the second most types of word-to-word combinations among all the texts reviewed in this chapter. It also has the most variety of sentence types. Its content is representative of mid-twentieth century New York City (with true proper nouns for people and places which no longer exist after more than half a century), but it has very animated content which could help speakers produce a variety of possible prosodic patterns.

One drawback of this passage is that it is long. Individual sentences also are long, and some of the words are rather difficult, although they are generally much easier than those in ERJ. It will probably place heavier burdens on the people reading it aloud in the recording session. Future research should investigate if its length may negatively affect the results.

3.3.9 Summary of Findings and Passage Selection

Naturally, the longer the passages are, the more phonetic coverage they have. Shorter passages are handy when it comes to recording, but they also have significant omissions. A major exception is the passage "The Boy who Cried Wolf" which has a relatively good phonetic coverage despite its shortness, but its sentence type coverage is not good enough probably because that was not a main concern when it was constructed. The longer passages under review can be tolerable. In fact, they have much fewer words than the sentence set used in my previous studies with *English Read by Japanese* database. Taking this into consideration, Labov's (2006) passage, which has the best score in each column, seems the strongest candidate for use in my study to collect read-aloud recordings of Japanese-accented English.

Table 1 displays a summary of the main characteristics of the passages surveyed in this chapter.

One possible objection to Labov's original text is that it contains only one instance of target /ʒ/. Also, I find it desirable that the text contain some instances of intervocalic voiced affricates which tend to be neutralized with fricatives in Japanese. So, I have decided to make the following minor adaptations:

Table 1 Summary of main characteristics in selected passages

	N of words	Phoneme coverage	*N* of consonant cluster types	*N* of word-to-word combination types	*N* of sentence types
"Stella"	69	Missing /ʒ, dʒ, j, aʊ, ɪɚ, ɛɚ, ɑɚ, ʊɚ/	23	33	4
"The North Wind and the Sun"	113	Missing /ʒ, ɑ, ɔɪ, ɪɚ, ɛɚ, ʊɚ/	20	45	2
"The Boy who Cried Wolf"	216	All	36	97	4
Prator and Robinett (1984)	165	Missing /ɪɚ, ɑɚ, ɔɪ/	32	45	8
Celce-Murcia et al. (2010)	226	Missing /ɔɪ/	42	75	9
Grant (2017)	138	Missing /ʒ, ɔɪ, ɛɚ, ɑɚ, ʊɚ/	32	66	5
"Arthur the Rat"	339	Missing /ʒ, ʊɚ/	40	119	5
Labov (2006)	349	All	44	110	10

- Well, we were waiting in (<— on) line about half an hour.
- "And what's the source of your information, Roger (<— Joseph)?"
- She suggested that he (<— told him to) ask a subway guard.
- Well, I managed to sleep through the worst part of the picture, and the stage show wasn't too hard to bear, which was a pleasure for me [inserted].

Although some of the lexical items are still a little difficult (e.g., Palisades, Paramount, rubies, etc.) and there are several words chosen specifically to elicit different characteristics of New York speech (e.g., "carry" vs. "Carey," "Mary" vs. "merry," "guard" vs. "God," "Chock" vs. "chalk"), I have made no effort to replace them with easier and more general ones. The adapted text (see Appendix) has 353 words (226 unique word forms) in the 9th-10th grade level.

4 Implications

Before conducting research for this chapter, Labov's (2006) passage was already my favorite candidate because of its animated content. Nevertheless, I was also wondering if it would be hard for speakers to handle because of some difficult lexical items and longer individual sentences. These could make the recording sessions more demanding. I now know that it is slightly better than others as far as the phonetic coverage is concerned.

The survey in this chapter has revealed that the phonetic coverage of the diagnostic passage is roughly in proportion to its length. While it is possible to make a passage that is more "efficient" for its shortness (like the "Wolf" passage), actually constructing one is quite another matter. The fact that many of the texts reviewed in this survey do not have a complete phonemic coverage reflects its sheer difficulty. All we can do is to look for some available materials before setting out to make a new passage, and to modify the one we have chosen so that it suits our own purposes more efficiently. I might want to refine Labov's text in some additional aspects, especially with respect to the coverage of word-to-word combinations, but then again, there may be no "ideal" passage fulfilling all the requirements. If I find it desirable to cover more patterns, the better idea will be to supplement the passage with a word list.

Another point to discuss is the uses of speech recordings collected with the passage. Fundamentally, we are not aware of all the pronunciation characteristics of Japanese speakers of English (or speakers of any L1, for that matter). The recordings help us not just to objectively confirm those that we know, but also to discover what has not been observed. Each of the passages analyzed in this chapter contains materials which can elicit the latter as well as many of the former. It is true that these short passages cannot uncover all the unknown pronunciation difficulties Japanese speakers of English (and other L1 speakers) may experience. That would require a passage or sentence set that contains all the patterns that can elicit them, and preparing such materials is unrealistic even if we do not set the upper limit to the number of words in it. As far as we do not require such an unrealistic comprehensiveness, however, collecting speech with passages is useful for phonetic studies. Also, these passages are useful in the teaching of pronunciation. They can be utilized not only for diagnostic purposes, as some of them were originally crafted for, but also as target models for learners.

5 Conclusion

In this chapter, I presented a linguistic analysis of eight different passages that can be used to assess L2 English speakers' read-aloud pronunciation. I proposed a set of requirements for the ideal passage, and analyzed the texts based on those requirements and with a focus on Japanese speakers of English. The passage that I have found meets most of the selection criteria is Labov's "text for phonemic contrast" (Labov, 2006). I slightly adapted the original text to make it better suited to elicit phonetic characteristics specific to Japanese speakers.

Acknowledgements The research for this study was supported by KAKEN Grant-in-Aid for Scientific Research (C) 18K00663 and (A) 18H04107 from the Japan Society for the Promotion of Science, and Chuo University Grant for Special Research.

Appendix

Labov's "text for phonemic contrasts" (Labov, 2006), with adapted part underlined.

Last Saturday night I took Mary Parker to the Paramount Theatre. I wanted to go and see The Jazz Singer, but Mary got her finger in the pie. She hates jazz, because she can't carry a tune, and besides, she never misses a new film with Cary Grant. Well, we were waiting <u>in</u> line about half an hour, when some farmer from Kansas or somewhere asked us how to get to Palisades Amusement Park.

Naturally, I told him to take a bus at the Port Authority Garage on 8th Avenue, but Mary right away said no, he should take the I.R.T. to 125th Street, and go down the escalator. She actually thought the ferry was still running.

"You're certainly in the dark," I told her. "They tore down that dock ten years ago, when you were in diapers."

"And what's the source of your information, <u>Roger</u>?" She used her sweet-and-sour tone of voice, like ketchup mixed with tomato sauce. "Are they running submarines to the Jersey shore?"

When Mary starts to sound humorous, that's bad: merry hell is sure to break loose. I remembered the verse from the Bible about a good woman being worth more than rubies, and I bared my teeth in some kind of a smile. "Don't tell this man any fairy tales about a ferry. He can't go that way."

"Oh yes he can!" she said. Just then a little old lady, as thin as my grandmother, came up shaking a tin can, and this farmer asked her the same question. She suggested that he ask a subway guard. My god! I thought, that's one sure way to get lost in New York City.

Well, I managed to sleep through the worst part of the picture, and the stage show wasn't too hard to bear, which was a pleasure for me. Then I wanted to go and have a bottle of beer, but she had to have a chocolate milk at Chock Full O' Nuts. Chalk this up as a total loss, I told myself. I bet that farmer is still wandering around looking for the 125th Street Ferry.

References

Cassidy, F. G. (Ed.). (1985). *Dictionary of American regional English* (Vol. 1). Harvard University Press.

Celce-Murcia, M., Brinton, D. M., Goodwin, J. M., & Griner, B. (2010). *Teaching pronunciation: A course book and reference guide* (2nd ed.). Cambridge University Press.

Deterding, D. (2006). The North Wind versus a Wolf: Short texts for the description and measurement of English pronunciation. *JIPA, 36*(2), 187–196.

Garofolo, J., Lamel, L., Fisher, W., Fiscus, J., Pallett, D., Dahlgren, N., & Zue, V. (1993). *TIMIT Acoustic-Phonetic Continuous Speech Corpus LDC93S1* (Electronic database). Linguistic Data Consortium. Retrieved January 29, 2022, from https://catalog.ldc.upenn.edu/LDC93S1

Grant, L. (2017). *Well Said: Pronunciation for clear communication* (4th ed.). National Geographic Learning.

Hiki, S., & Kakita, K. (2013). The 'panphonic' text of 'The North Wind and the Sun' for the illustration of the International Phonetic Alphabet of Japanese consonants and its use in the phonetic analysis of Japanese speech. *Proceedings of Meetings on Acoustics*, *19*, 060285. ASA. https://doi.org/10.1121/1.4800255

The International Phonetic Association. (1999). *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge University Press.

Kominek, J., & Black, A. W. (2003). CMU ARCTIC: Databases for speech synthesis. Online document. *Carnegie Mellon University Language Technologies Institute Tech Report* CMU-LTI-03–177. Retrieved January 29, 2022, from http://festvox.org/cmu_arctic/cmu_arctic_report.pdf

Labov, W. (2006). *The social stratification of English in New York City* (2nd ed.). Cambridge University Press.

Ladefoged, P., & Johnson, K. (2015). *A course in phonetics* (7th ed.). Cengage Learning.

Maier, P. (2019). *IDEA international dialects of English archive.* Retrieved January 29, 2022, from https://www.dialectsarchive.com/

Makino, T. (2013). Pronunciation characteristics of Japanese speakers' English: A preliminary corpus-based study. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 5th Pronunciation in Second Language Learning and Teaching Conference* (pp. 120–136). Iowa State University. Retrieved January 29, 2022, from https://apling.engl.iastate.edu/wp-content/uploads/sites/221/2015/05/PSLLT_5th_Proceedings_2013.pdf

Makino, T., & Aoki, R. (2012). English read by Japanese Phonetic Corpus: An interim report. *Research in Language, 10*(1), 79–95. https://doi.org/10.2478/v10015-011-0046-5

Meng, H., Tseng, C.-Y., Kondo, M., Harrison A., & Viscelgia, T. (2009). Studying L2 suprasegmental features in Asian Englishes: A position paper. In R. Moore (Ed.), *Proceedings of Interspeech 2009* (pp. 1715–1718). ISCA. https://doi.org/10.21437/Interspeech.2009-517

Minematsu, N., Okabe, K., Ogaki, K., & Hirose, K. (2011). Measurement of objective intelligibility of Japanese accented English using ERJ (English Read by Japanese) database. In P. Cosi & R. De Mori (Eds.), *Proceedings of Interspeech 2011* (pp. 1481–1484). ISCA. https://doi.org/10.21437/Interspeech.2011-310

Minematsu, N., Tomiyama, Y., Yoshimoto, K., Shimizu, K., Nakagawa, S., Dantsuji, M., & Makino, S. (2002). English speech database read by Japanese learners for CALL system development. *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC 2002)* (pp. 896–903). European Language Resources Association. Retrieved January 29, 2022, from http://www.lrec-conf.org/proceedings/lrec2002/pdf/155.pdf

Prator, C. H., & Robinett, B. W. (1984). *Manual of American English pronunciation* (4th ed.). Harcourt College Publishers.

Ueyama, M. (1997). The phonology and phonetics of second language intonation: the case of "Japanese English". In G. Kokkinakis (Ed.), *Proceedings of the 5th European Conference on Speech Communication and Technology (Eurospeech 1997)* (pp. 2411–2414). ISCA. Retrieved January 29, 2022, from https://www.isca-speech.org/archive/pdfs/eurospeech_1997/ueyama97_eurospeech.pdf

Vance, T. (2008). *The sounds of Japanese.* Cambridge University Press.

Venditti, J. J. (2005). The J_ToBI model of Japanese intonation. In S.-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199249633.003.0007

Weinberger, S. (2015). *Speech accent archive.* George Mason University. Retrieved January 29, 2022, from http://accent.gmu.edu

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

Zhao, G., Sonsaat, S., Silpachai, A., Lucic, I., Chukharev-Hudilainen, E., Levis, J., & Gutierrez-Osuna, R. (2018). L2-ARCTIC: A non-native English speech corpus. In B. Yegnanarayana (Ed.), *Proceedings of Interspeech 2018* (pp. 2783–87). ISCA. https://doi.org/10.21437/Interspeech.2018-1110

Takehiko Makino is Professor of English as a Foreign Language at Chuo University in Tokyo, Japan. He was awarded a Certificate of Proficiency in the Phonetics of English by IPA in 2004. His publications include the Japanese translation of Peter Ladefoged's *A Course in Phonetics* (1999) and an original textbook (in Japanese) *English Phonetics for Japanese Speakers* (2005).

Speech Perception

Accentedness and Comprehensibility in Non-native Listeners' Perception of L2 Speech

Joan C. Mora

Abstract This study examined, from an individual differences perspective, the relationship between accentedness and comprehensibility in non-native English for non-native listeners. Forty non-native learners of English differing in L1 (20 L1-Catalan; 20 L1-German) and L2 proficiency level (10 low, 10 high within each L1 group) and 10 native English speakers performed two 60-trial rating tasks based on two 7-point Likert scales, one for accentedness and one for comprehensibility. The sentence stimuli were 10 different true/false English sentences spoken by four non-native English learners at an intermediate proficiency level (two L1-Catalan and two L1-German) and two native English speakers, so that each listener rated the same sentences six times, two in each accent (Catalan-accented, German-accented and native English). Non-native listeners perceived sentences spoken by L1-matched speakers as more weakly accented and comprehensible than those spoken by L1-unmatched speakers, irrespective of L2 proficiency level. However, all sentences were judged to be less comprehensible by low- than high-proficiency listeners, and high- but not low-proficiency listeners found Catalan- and German-accented sentences as comprehensible as native listeners did. Analyses of individual listener data revealed that inter-listener variation in how strongly accentedness was related to comprehensibility was dependent on non-native listeners' L1 background and L2 proficiency level.

Keywords Accentedness · Comprehensibility · Non-native speech · Non-native listeners · L2 proficiency

---

J. C. Mora (✉)
Faculty of Philology and Communication, University of Barcelona, Barcelona, Spain
e-mail: mora@ub.edu

V. G. Sardegna and A. Jarosz (eds.), *Theoretical and Practical Developments in English Speech Assessment, Research, and Training*, Second Language Learning and Teaching,
https://doi.org/10.1007/978-3-030-98218-8_7

1 Introduction

Non-native spoken communication in English most often takes place between speakers of different L1 backgrounds and English proficiency levels (Pennycook, 2017). Fluent communication between non-native interlocutors largely depends on their ability to understand others and make themselves understood despite the presence of unfamiliar accents that may be detrimental to intelligibility (Bent & Bradlow, 2003) and differing levels of L2 competence that may pose a threat to comprehensibility. In line with this reality, and supported by decades of research on accentedness, intelligibility, and comprehensibility (e.g., Munro & Derwing, 1995a; Saito et al., 2016a), new developments in L2 pronunciation teaching and learning have witnessed a shift of focus from nativelikeness and accentedness to speech comprehensibility (Derwing & Munro, 2015; Isaacs, 2018; Levis, 2005). However, most research investigating the linguistic correlates of comprehensibility (Isaacs & Trofimovich, 2012; Saito et al., 2016a, 2016b), as well as much of the research investigating the relationship between foreign accent, comprehensibility, and perceived fluency (Derwing & Munro, 1997; Munro & Derwing, 1995a; Pinget et al., 2014) has primarily relied on native speakers' perceptual judgements of non-native speech, rather than on non-native speakers' perceptual judgements. The current study takes an individual differences approach in examining the relationship between accentedness and comprehensibility in non-native English from the perspective of non-native listeners by exploring the extent to which this relationship is modulated by L2 listener characteristics (L1 background and proficiency level).

2 Literature Review

2.1 Factors Modulating Perceptual Judgements of Non-Native Speech

Accentedness and comprehensibility are two of the main perceptual dimensions of non-native speech examined in relation to L2 learners' oral production ability. They are related to, and partly independent from, intelligibility—that is, the extent to which a spoken utterance is understood by the listener (Kennedy & Trofimovich, 2008; Munro & Derwing, 1995a)—and fluency, speech smoothness and fluidity (Lennon, 2000; O'Brien, 2014). Accentedness refers to how closely the pronunciation of an utterance matches that of a native speaker (Kennedy & Trofimovich, 2008), whereas comprehensibility refers to listeners' perception of how easy or difficult it is for them to understand L2 speech (Derwing et al., 2008). Both dimensions are typically assessed through scalar judgements (e.g., 9-point Likert scales) of 20/30 s-long L2 speech samples elicited through picture-based monologic oral narrative tasks (see Thomson, 2018, for an overview of measurement methods). Non-native (accented) speech is generally perceived to be less fluent, less intelligible, and less comprehensible than native speech and has been shown to slow down processing (Ludwig & Mora, 2017; Munro & Derwing, 1995b). Accentedness and comprehensibility can be assessed as independent constructs, as it is perfectly possible for non-native speakers to speak with a strong accent and still be understood without difficulty (Munro & Derwing, 1995a; Munro et al., 2006). Speech features (i.e., linguistic properties of the speech samples), listeners' characteristics, and even the kind of speaking tasks used to elicit L2 speech (Crowther et al., 2015a) may affect perceptual assessments of L2 speech.

Recent research on the linguistic correlates of accentedness and comprehensibility has shown that L2 learners' phonetic and phonological speech features that differ from those of native speakers (e.g., inaccurate realization of L2 speech sounds, phonemic substitutions, misplacement of lexical stress, prosodic appropriateness) contribute most strongly to L1 listeners' perceived degree of accentedness. In contrast, the linguistic features of L2 speech that contribute to L1 listeners' perception of degree of comprehensibility (i.e., the amount of effort listeners need to put into understanding L2 speech) include, besides phonetic and phonological features, a variety of time-based fluency phenomena (speech rate, articulation rate, pause frequency, location and duration) as well as lexical and grammatical accuracy, richness and complexity (Crowther et al., 2015b; Derwing & Munro, 1997; Isaacs & Trofimovich, 2012; Munro & Derwing, 1995a; Saito et al., 2016b; Saito et al., 2017; Trofimovich & Baker, 2006). The relationship between listeners' perception of accent and comprehensibility may therefore be partly explained by the fact that accent is one of the features of non-native speech that may impact the amount of effort listeners need to put into understanding non-native utterances.

Listeners' characteristics, such as their experience or familiarity with a specific L2 accent, or their L1 background (i.e., the extent to which the listeners' L1 and the speakers' L2 are typologically related or the degree of L1-L2 mutual intelligibility), may affect their accentedness and comprehensibility judgements (Crowther et al., 2015b; Munro et al., 2006), but research to date has produced somewhat mixed results. For example, Kennedy and Trofimovich (2008) found native listeners differing in experience (extent of previous exposure to non-native speech) to rate accentedness and comprehensibility similarly, and Derwing and Munro (2013) found native and high-proficiency non-native speakers of English not to differ in how they rated L2 speech for these dimensions. However, research has also shown that raters are more lenient when they are familiar with the type of accent they are asked to judge (Foote & Trofimovich, 2018; Winke et al., 2013), exhibiting a processing advantage for non-native speech in their own L2 accent (Ludwig & Mora, 2017).

2.2 Relationship Between Accentedness and Comprehensibility

Although intelligibility, fluency, accentedness, and comprehensibility are related to one another, the degrees and the strength of these associations vary from study to study (Thomson, 2018), accentedness and comprehensibility are often reported to be as strongly related to one another as comprehensibility and intelligibility, whereas accentedness is more weakly related to intelligibility (Jułkowska & Cebrian, 2015; Munro & Derwing, 1995a). Scalar judgements (e.g., 9-point Likert scales) of these dimensions, however, reveal differences in the degree of accentedness and comprehensibility listeners perceive in non-native speech. Studies assessing the accentedness and comprehensibility in the same set of speech samples have consistently shown that the proportion of high comprehensibility ratings (ease or little difficulty in understanding) is much higher than the proportion of low accentedness ratings (little or no accent), irrespective of whether the speech samples assessed by native speakers consisted of picture-elicited oral narratives (Derwing & Munro, 1997) or isolated sentences obtained through a delayed sentence repetition task (Mora & Darcy, 2016) or whether the speech samples were assessed by non-native listeners (Jułkowska & Cebrian, 2015). That is, listeners perceive much higher levels of accentedness than they do of difficulty in understanding, which underscores the relative independence of accentedness and comprehensibility as perceptual dimensions of non-native speech.

There is a dearth of research on the relationship between accentedness and comprehensibility from the perspective of non-native listeners. Most research investigating the perception of non-native speech by non-native listeners has focused on intelligibility and potential L1-matched and unmatched speech intelligibility benefits (Bent & Bradlow, 2003; Hayes-Harb et al., 2008; Stibbard & Lee, 2006), whereas studies investigating accentedness and comprehensibility by non-native listeners have been mainly concerned with identifying the differential weight various speech features have on these dimensions (Crowther et al., 2016; O'Brien, 2014; Saito et al., 2019). Previous studies have shown a relatively strong association between accentedness and comprehensibility ratings, but they all report large variability among listeners. For example, Munro and Derwing (1995a) found accent and comprehensibility scores for Mandarin talkers of English performing an oral narrative task to correlate significantly for 17 of their 18 listeners, but the strength of the correlation ranged from $r = 0.41$ to $r = 0.82$. Similarly, in Munro and Derwing (1995b), the relationship between accent and comprehensibility for Mandarin talkers of English producing short sentences reached a significant *Pearson-r* correlation coefficient of $r = 0.624$, but for six out of the 20 native listeners the correlation was non-significant, and in fact correlations varied greatly in strength (from $r = 0.140$ to $r = 0.917$). The one study examining the relationship between accentedness and comprehensibility in non-native listeners' ratings of L2 speech (Polish-accented English) we are aware of (Jułkowska & Cebrian, 2015) found significant correlation coefficients that varied in strength as a function of listeners' L1: English ($n = 6$; $r = 0.804$), Polish ($n = 6$; $r = 0.344$) and Spanish ($n = 6$; $r = 0.557$), but only 10 out of the 18 listeners' correlations reached significance (variability in the strength of correlations within listener groups is not reported). For the homogeneous L1 listener groups in Munro and Derwing's (1995a, 1995b) studies, variability in how strongly accentedness relates to comprehensibility suggests that some listeners paid attention to accent when judging ease of understanding, while others did not. Other kinds of individual differences, such as memory and attention (Isaacs & Trofimovich, 2011) or awareness of the importance of accent and comprehensibility for communication (Saito et al., 2020), might be at play, too. For the non-native listener groups in Jułkowska and Cebrian (2015), between group differences were attributed to listeners' L1 and accent familiarity as the researchers interpreted the weak correlation in the Polish listeners' group in terms of their better ability to understand Polish-accented English irrespective of degree of accentedness. These findings seem to suggest that the influence of accent on comprehensibility may be of a larger magnitude for listeners whose phonology differs the most from that of the speech input (i.e., native speakers). In addition, listeners' proficiency may not only affect accent (Eger & Reinisch, 2019) and comprehensibility ratings (Saito et al., 2019), but also how these dimensions relate to one another. The current study extends this line of research by assessing the relationship between accentedness and comprehensibility in non-native listeners differing in L1-background (which may or may not match the non-native speakers' accent) and L2 proficiency.

3 The Study

This study examined the relationship between the accentedness and comprehensibility ratings of non-native English for 40 non-native listeners (L2-English learners) differing in L1 background and L2 proficiency level. In a previous study based on data from the same participant pool (Ludwig & Mora, 2017), we explored the relationship between listeners' processing times and comprehensibility judgments and found that processing costs in sentence comprehension were associated with perceived effort in understanding, but this relationship, which was revealed through significant moderate correlations, was mediated by an interaction between listeners' L1 and their L2 proficiency. That is, accented English was processed faster and judged to be more comprehensible by non-native listeners if produced by L1-matched speakers, and it was faster to process and easier to understand than native English by low-proficiency listeners, whereas high-proficiency listeners showed a processing advantage over native English listeners. The present study extends these analyses by including the accentedness ratings provided by the same listeners on the same speech samples previously judged for comprehensibility, and by focusing on the relationship between accentedness and comprehensibility at an individual listener level. Our aim was to explore individual differences in non-native listeners' judgments of accentedness and comprehensibility in L2 speech as a function of L1 background and L2 proficiency. We therefore addressed the following two research questions (RQ):

RQ1: Are non-native listeners' ratings of accentedness and comprehensibility equally affected by L1 background (matched- vs. mismatched-L1) and proficiency (low vs. high)?

RQ2: Does L1 background and L2 proficiency modulate the relationship between accentedness and comprehensibility in non-native listeners?

4 Methods

4.1 Participants

4.1.1 Speakers

Two Catalan learners of English (1 female, 1 male), Catalan-dominant Catalan-Spanish bilinguals born and raised in Catalonia (Spain); and two German learners of English (1 female, 1 male), born and raised in Germany, were selected as non-native speakers of English from a larger pool of upper-intermediate EFL learners. Two native English speakers (1 female, 1 male), born and raised in the United Kingdom, were selected as native speakers. They were all selected on the basis of clarity of articulation and absence of pronunciation errors and hesitations (aged 21–25). The Catalan and German speakers had never lived in an English-speaking country and had learnt English as adults through formal instruction in a foreign language context. Their self-reported level of proficiency in English ranged from 3 to 4 on a 5-point Likert scale (1 = elementary; 5 = near-native). They read sentences from randomized printed lists that were recorded, spliced and normalized for amplitude to be used as speech stimuli. A one-way ANOVA showed that the three speaker groups produced the sentence stimuli at similar articulation rates: $F(2,117) = 1.29$, $p = 0.277$.

4.1.2 Listeners

The listeners were 20 native speakers of Catalan, 20 native speakers of German and 10 control native speakers of English. They were born and raised in Catalonia (Spain), Germany, and the United Kingdom, respectively. The non-native listeners had never lived outside their home country and were unfamiliar with non-native English accents other than their own. They were recruited at language schools from intermediate- and advanced-level groups (10 of each within the Catalan and German speaker groups). A vocabulary size measure confirmed listener groups had non-overlapping distributions (Table 1).

Table 1 Listeners' characteristics as a function of L1 and L2 proficiency (*SD*s in parentheses)

Listeners	Catalan ($n = 20$)		German ($n = 20$)		English ($n = 10$)
Age (years)	24.2 (1.06)		24.4 (1.31)		24.3 (1.19)
LoR (years)[a]	23.5 (2.07)		21.75 (2.05)		22.8 (2.03)
L2 proficiency	High	Low	High	Low	Native
FI (years)[b]	7 (0.84)	6.2 (1.03)	9.2 (0.63)	9 (0)	–
Proficiency (1–5)[c]	4.1 (0.57)	2.3 (0.48)	4.5 (0.53)	2.5 (0.71)	5 (0)
Vocabulary size (0–10,000)	6620 (481.4)	3215 (189.7)	6685 (460.7)	3300 (143.37)	9750 (156.4)
Fam-Cat (1–5)[d]	4.7 (0.48)	4.9 (0.32)	1.3 (0.48)	1.1 (0.32)	1.6 (0.39)
Fam-Ger (1–5)[e]	1.2 (0.42)	1 (1)	4.8 (0.42)	4.7 (0.48)	1.5 (0.79)

[a]LoR = length of residence in home country
[b]FI = years of formal instruction in English
[c]Proficiency = mean self-rated proficiency
[d]Fam-Cat = mean familiarity with Catalan-accented English
[e]Fam-Ger = mean familiarity with German-accented English

4.2 Materials, Rating Tasks, and Procedures

The elicited sentences were based on the single-clause statements in Munro and Derwing's (1995b) sentence verification task. Sixty different sentences (10 by each speaker) were recorded in a sound-proof booth, normalized for peak and mean amplitude and high-pass filtered (50 Hz). Cross-language cognate status was controlled for and content words were selected from within the 2000 most frequent English words. Sentences were comparable across accents in word length ($M = 5.66$, $SD = 5.6$; $F(2,57) = 0.617$, $p = 0.543$), syllable length ($M = 8.38$, $SD = 1.4$; $F(2,57) = 0.610$, $p = 0.547$), speech rate in syllables per second ($M = 0.47$, $SD = 0.51$; $F(2,57) = 1.909$, $p = 0.1580$) and duration ($M = 2421$, $SD = 318$; $F(2,57) = 0.374$, $p = 0.690$).

The 60 sentences were presented randomly to listeners twice in two separate computer-administered rating tasks, one for accentedness and one for comprehensibility. Listeners rated accentedness and comprehensibility on 7-point Likert scales (*1 = No foreign accent, 7 = Very strong foreign accent; 1 = Very easy to understand, 7 = Very difficult to understand*). In the comprehensibility rating task, the sentences were embedded in cafeteria noise (SNR = 10 dB) to help listeners focus on perceived difficulty in understanding. Listeners were explained the difference between accentedness and comprehensibility and were encouraged to use the whole scale. They could listen to every sentence maximally twice before making a decision. Task instructions were given in their L1 and tasks were performed individually in one 45-min session.

5 Results

5.1 *Listeners' L1 and Proficiency Effects on Accentedness and Comprehensibility*

Listeners' accentedness and comprehensibility ratings were consistent among listeners within listener groups (intra-class correlation coefficients $\alpha > 0.9$), thus indicating homogeneity of ratings. The averaged ratings (Table 2) show differences between accentedness and comprehensibility as well as differences as a function of listeners' L1 and L2 proficiency.

For sentences spoken in a non-native accent, ratings were overall higher for accentedness than for comprehensibility (see Fig. 1), in line with previous research findings (Jułkowska & Cebrian, 2015; Munro & Derwing, 1995a). Catalan-accented sentences were consistently perceived by all listener groups to be more strongly accented than German-accented sentences, even by L1-Catalan listeners. This suggests that a Catalan accent might be perceived by all listeners as being more distant from native

Table 2 Mean accentedness and comprehensibility ratings (0–7) as function of listeners' L1 and proficiency level and sentence accents (standard deviations in parentheses)

	Accentedness			Comprehensibility		
	Catalan	German	English	Catalan	German	English
L1-Catalan	5.63 (0.47)	4.89 (0.48)	1.28 (0.39)	4.25 (0.86)	5.00 (0.63)	2.07 (0.71)
Low	5.31 (0.35)	4.71 (0.42)	1.40 (0.44)	3.80 (0.58)	5.19 (0.47)	2.56 (0.59)
High	5.96 (0.33)	5.08 (0.49)	1.17 (0.31)	4.70 (0.88)	4.82 (0.74)	1.58 (0.42)
L1-German	6.04 (0.47)	4.74 (0.51)	1.32 (0.42)	5.24 (0.66)	3.95 (0.58)	2.16 (0,76)
Low	6.04 (0.46)	4.56 (0.45)	1.42 (0.42)	5.69 (0.26)	4.03 (0.54)	2.77 (0.37)
High	6.05 (0.49)	4.92 (0.53)	1.22 (0.41)	4.80 (0.63)	3.88 (0.63)	1.56 (0,52)
L1-English	6.16 (0.25)	5.06 (0.54)	1.05 (0.15)	4.15 (0.55)	3.50 (0.44)	1.23 (0.31)

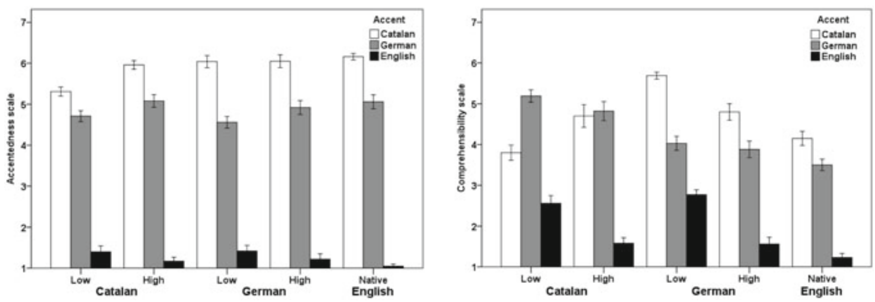

Fig. 1 Mean accentedness (left) and comprehensibility (right) ratings according to listeners' L1 and L2 proficiency (low, high, native) (error bars = $\pm$ 1*SE*)

English than a German accent. Neither listeners' L1 background nor L2 proficiency seem to have affected accentedness ratings much (Fig. 1 left), whereas they seem to have had a large effect on comprehensibility (Fig. 1 right). A $2 \times 2 \times 2$ ANOVA on accentedness ratings with non-native listeners' *L1* (Catalan, German) and *Proficiency* (low, high) as between-subjects factors and sentence *Accent* (Catalan, German, native English) as the within-subjects factor, revealed a significant main effect of *Accent* ($F(2, 44) = 3722.8$, $p < 0.001$, $\eta^2 = 0.994$), but neither the effect of *L1* ($F(1, 45) = 0.87$, $p = 0.353$, $\eta^2 = 0.019$) nor *Proficiency* ($F(1, 45) = 2.41$, $p = 0.128$, $\eta^2 = 0.051$) reached significance. Bonferroni-adjusted pairwise comparisons showed that listeners perceived a significantly stronger accent on Catalan-accented sentences than on German-accented sentences ($p < 0.001$), both of which were obviously perceived as significantly more strongly accented than sentences spoken in a native English accent. However, *Accent* significantly interacted with *Proficiency* ($F(2, 44) = 11.83$, $p < 0.001$, $\eta^2 = 0.350$) and *L1* and *Proficiency* ($F(2, 44) = 4.62$, $p = 0.015$, $\eta^2 = 0.174$) because, according to Tukey post-hoc tests, Catalan-accented sentences were perceived to be significantly less strongly accented by L1-Catalan than by L1-German listeners ($p = 0.013$) or native English listeners ($p = 0.009$). This suggests that an L1 match between listener and speaker may result in more lenient accentedness ratings, that is, L1-Catalan listeners perceived less of an accent in Catalan sentences than L1-German listeners did (and vice-versa, though not significantly).

The comprehensibility ratings were submitted to the same $2 \times 2 \times 3$ ANOVA just described, revealing significant main effects of *Proficiency* ($F(1, 45) = 18.96$, $p < 0.001$, $\eta^2 = 0.296$) and *Accent* ($F(2, 44) = 352.13$, $p = 0.001$, $\eta^2 = 0.941$), but no main effect of *L1* ($F(1, 45) = 0.17$, $p = 0.898$, $\eta^2 < 0.001$). However, a complex set of significant interactions arose (*L1* × *Accent*, *L1* × *Proficiency*, *Proficiency* × *Accent*, *L1* × *Proficiency* × *Accent*), suggesting that non-native listeners' comprehensibility ratings were affected by a match between their L1s and the sentence accents as well as their L2 proficiency level. In fact, all listener groups were found to rate all accents differently for comprehensibility (all $p < 0.001$). Accented sentences were significantly more comprehensible when the listeners' and speakers' L1 matched, whereas native English sentences were found to be equally comprehensible for L1-Catalan and L1-German listeners ($p = 0.898$). As regards proficiency, native English sentences were less comprehensible for low- than for high-proficiency listeners, as expected (all $p < 0.001$), whereas Catalan- and German-accented sentences were equally comprehensible for high-proficiency and native English listeners (all $p > 0.1$). However, all sentences were harder to understand for low- than for high-proficiency listeners irrespective of listeners' L1 (all $p < 0.001$).

To sum up, non-native listeners found sentences spoken in their own accent to be less strongly accented and easier to understand than those spoken in an unfamiliar accent, whereas L1-English listeners found all accented utterances to have a similar level of accentedness and to be equally difficult to understand. Interestingly, native English sentences were perceived to be the easiest to understand by all listener groups.

5.2 Relationship Between Listeners' Ratings of Accentedness and Comprehensibility

To explore how listeners' ratings for accentedness and comprehensibility related to one another, we conducted three sets of analyses. In these analyses, we included Catalan- and German-accented sentences ($n = 40$) only, as sentences spoken in a native English accent had on average received accentedness ratings of 1 (i.e., they were perceived as having no accent). First, we ran *Pearson-r* correlations on all ratings as a function of listeners' L1 and L2 proficiency (400 ratings per subject group: 10 raters × 40 sentences, 20 in each accent). These analyses revealed significant positive correlations between the accentedness and comprehensibility ratings, except for the ratings from the low-proficiency Catalan listeners ($r = 0.276$, $p < 0.001$) for whom a stronger accent in the speech samples appeared to be weakly associated with ease of understanding. The correlation coefficients that resulted from the ratings by high-proficiency Catalan listeners ($r = 0.128$, $p = 0.01$), and those of the German (*low proficiency*: $r = 0.530$, $p < 0.001$; *high proficiency*: $r = 0.272$, $p < 0.001$) and English ($r = 0.373$, $p < 0.001$) listeners were all positive and weak-to-moderate in strength.

Second, we explored group correlations (10 listeners per group) between accentedness and comprehensibility as a function of listeners' L1 and L2 proficiency based on the averaged 20 ratings each listener provided per accent. As shown in Fig. 2, group differences in how accentedness was related to comprehensibility for non-native listeners mainly concerned low proficiency listeners, for whom there was a comprehensibility benefit in their own accent (i.e., less difficulty in understanding for speech in their own accent), whereas high-proficiency listeners showed larger overlap in the ratings for Catalan- and German-accented sentences. Given the low number of participants per group ($n = 10$) none of the correlations plotted below reached statistical significance.

Finally, we computed individual *Pearson-r* correlations by listener based on the ratings for accentedness and comprehensibility for each of the 40 sentences rated. These analyses revealed large individual variability in both the strength and the direction of the correlations. For example, some of the correlations for L1-Catalan listeners were negative and some were non-significant. Although for a majority of listeners accentedness was positively and significantly related to comprehensibility, the strength and direction of this relationship varied as a function of the listeners' L1 and L2 proficiency level. As shown in Table 3, the number of listeners for whom accentedness was significantly and positively related to comprehensibility varied as a function of subject group (Fig. 3). For the German and native English listeners, accentedness was always positively associated with comprehensibility—that is, a stronger accent was associated with greater difficulty in understanding –, whereas most Catalan listeners, especially low-proficiency listeners, perceived more strongly accented sentences to be easier to understand, thus showing a comprehensibility benefit for non-native speech. In general, these results support previous findings on the relationship between accentedness and comprehensibility for native (Munro & Derwing, 1995a) and non-native (Jułkowska & Cebrian, 2015) listeners.

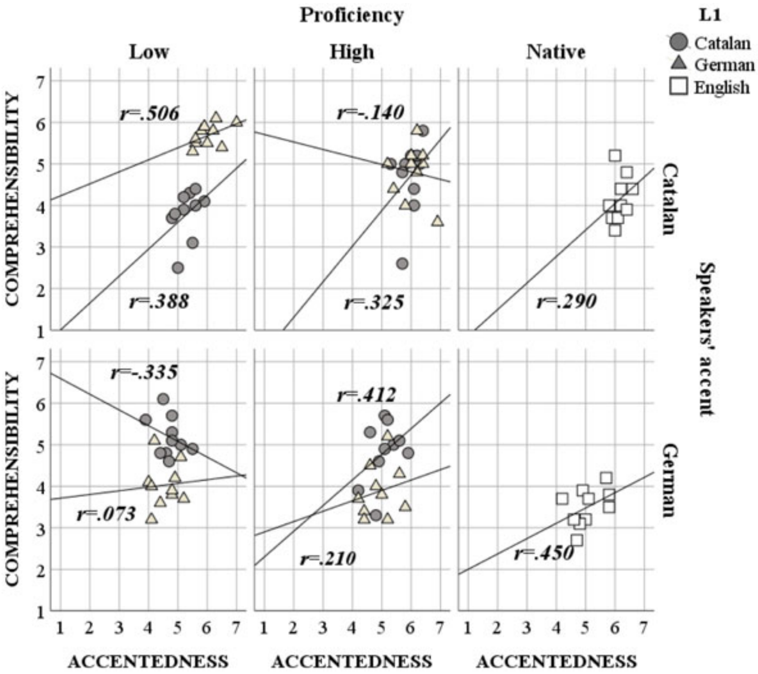

Fig. 2 *Pearson-r* correlations between accentedness and comprehensibility as a function of listeners' L1 and L2 proficiency

Table 3 Number of listeners (out of 10 per group) who obtained positive (+), negative (–), positive significant (+Sig) and significant (Sig) correlations

Proficiency	Low				High				Native			
	+	–	+*Sig*	*Sig*	+	–	+*Sig*	*Sig*	+	–	+*Sig*	*Sig*
L1												
Catalan	1	9	1	6	5	5	4	5	–	–	–	–
German	10	1	8	8	10	0	5	5	–	–	–	–
English	–	–	–	–	–	–	–	–	10	0	5	5

6 Discussion

The present study confirms and extends previous findings by Munro and Derwing (1995a) as well as Jułkowska and Cebrian (2015) on the relationship between accentedness and comprehensibility in several ways. Unlike Munro and Derwing (1995a), we focused on the perception of accentedness and comprehensibility by non-native listeners, as Jułkowska and Cebrian (2015) did. We also followed up on Jułkowska and Cebrian's study by including both a match and a mismatch between the accent

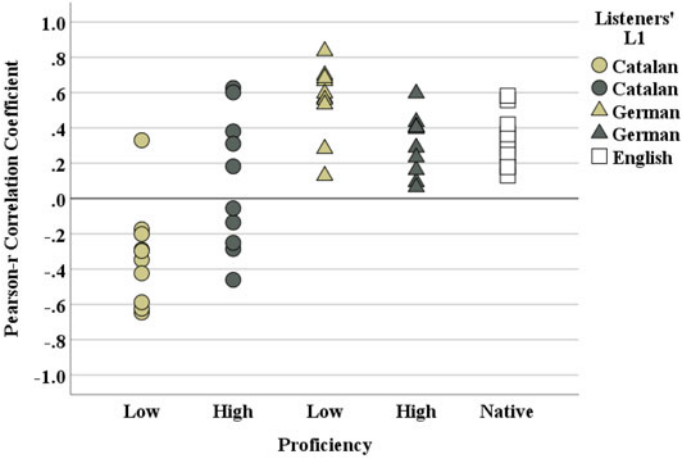

Fig. 3 Individual *Pearson-r* correlations between accentedness and comprehensibility as a function of listeners' L1 and L2 proficiency

in the speech samples and listeners' L1 for *two* listeners' L1s (Catalan- and German-accented sentences presented to L1-Catalan and L1-German listeners) rather than *one* (Polish-accented speech presented to L1-Polish and L1-Spanish listeners). We were able to show not only L1-based differences in how non-native listeners rate speech samples for accentedness and comprehensibility due to the presence or absence of a match between accent and L1, but also L1-based differences in how accentedness was related to comprehensibility for L1-matched sentences. Namely, whereas most L1-Catalan listeners (low-proficiency listeners in particular) perceived more strongly accented sentences in a Catalan accent to be easier to understand, this was not the case for the L1-German listeners, who generally showed a positive relationship between degree of accent and difficulty in understanding. These findings indicate that L1 background differences may modulate the relationship between accentedness and comprehensibility. In addition, we also extended Jułkowska and Cebrian's research by exploring the effects of non-native listeners' proficiency on the perception of accent-matched and mismatched sentences and on the relationship between accentedness and comprehensibility.

With respect to accent-based differences in the perception of non-native speech, the present study shows (confirming previous findings) that non-native listeners perceive speech in their own accent to be slightly less strongly accented than speech in an unfamiliar accent, despite overall differences in degree of accentedness between accents. This finding suggests that exposure to an accent leads to more lenient ratings, indicating a certain level of accent sensitivity loss. This effect was modulated by L2 proficiency: high-proficiency listeners perceived a stronger degree of accent than low-proficiency listeners did, and it appeared to be independent from listeners' L1, as Catalan-accented sentences were always perceived by all listeners to be more strongly accented than German-accented sentences. This difference might be explained by the Catalan speakers being less proficient than the German speakers, or by a Catalan accent being more distant from native English than a German accent, or both. This finding lends support to previous research findings (e.g., Jułkowska & Cebrian, 2015; Munro et al., 2006) indicating that accentedness judgements are relatively independent from listeners' L1 background and L2 proficiency.

As regards comprehensibility ratings, two relevant outcomes were obtained. First, the data showed a robust interlanguage speech comprehensibility benefit, so that non-native listeners found sentences in their own accent easier to understand than sentences in an unfamiliar accent (irrespective of listeners' L1). However, this effect was found to be modulated by listeners' proficiency, as the size of the effect was strong for low-proficiency learners, but diminished for high-proficiency German listeners and disappeared for high-proficiency Catalan listeners. Second, sentences spoken in a native English accent were judged to be easier to understand than sentences spoken in either a familiar or unfamiliar non-native accent for both low- and high-proficiency listeners of both L1 backgrounds. This result might be due to listeners' greater exposure to L2 input in native English (e.g., through media) than in familiar or unfamiliar non-native accents, the lack of disfluencies and lexical and grammatical inaccuracies typical of more spontaneous types of speech (e.g., oral narratives), or the possibility that the non-native listeners paid attention to degree of accentedness when rating the speech samples for comprehensibility. Further research exploring the effects of these factors on comprehensibility ratings is warranted.

The individual data analyses on the relationship between accentedness and comprehensibility were generally in accordance with previous findings for both native (Munro & Derwing, 1995a, 1995b) and non-native (Jułkowska & Cebrian, 2015) listeners, but revealed large variability in the strength and direction of the relationship between the two dimensions. Although in general the relationship was significant and positive (i.e., listeners had a greater difficulty in understanding more accented speech), the majority of L1-Catalan listeners, especially those with low proficiency, associated stronger accentedness to easiness (rather than difficulty) in understanding. Comprehensibility benefits of an L1-matched accent (also present in L1 German listeners) might explain this outcome. It is also possible that German-accented sentences were easier to understand by L1-Catalan listeners than Catalan-accented sentences for L1-German listeners because the German accent is closer (than Catalan) to native English—that is, the L2 of the listeners. These findings underscore the potential effects of L1 background (an L1-match between listener and speaker, and closeness to the L2) and proficiency level in non-native listeners' perception of L2 speech as they both appear to modulate how accentedness relates to comprehensibility. Further research with other L1s is needed to corroborate the findings of this exploratory study.

7 Implications

The outcomes of the present study suggest a number of implications for L2 pronunciation teaching, assessment, and research. In L2 pronunciation teaching, a common recommendation (for the average L2 learner) is to focus on effective communicative competence, and consequently on those dimensions of speaking performance that make L2 learners' speech detrimental to comprehensibility, rather than on nativelikeness and pronunciation accuracy (Darcy, 2018; Derwing & Munro, 2005). In foreign language teaching contexts, learners are often exposed to L1-accented speech (from peer students or their teachers, or both), for which the current study shows benefits in comprehensibility. However, high levels of comprehensibility achieved on the basis of the common phonetic substrate of the listener's L1 and L1-accented L2 speech may be deceiving for learners, resulting in a comprehensibility cost when exposed to unfamiliar non-native accents of English. As shown in the present study, a non-native unfamiliar accent was detrimental to comprehensibility, especially to low-proficiency listeners, whereas native English was not. Thus, from a pedagogical perspective, it would seem convenient to expose learners to non-native accents other than their own besides exposing them to L2 speech by native speakers (Derwing et al., 2002).

Accentedness and comprehensibility are two important dimensions of L2 pronunciation assessment (Kang & Ginther, 2018). The present study has shown that non-native listeners' assessment of these dimensions is partly dependent on listeners' L1 background, L2 proficiency level, and how L2 speech is produced by listeners, which may determine familiarity with a specific accent. When assessing L2 pronunciation, therefore, instructors and testers should be aware of the potential biases that may affect their judgements resulting from L1-matched and mismatched L2 speech.

Finally, the current study has shown large variability in how non-native listeners relate accentedness to comprehensibility, partly modulated by L1 and proficiency effects. Whereas for some listeners these two dimensions appear to be completely independent from one another, for others they are strongly and positively associated. This largely under-researched area in L2 speech studies deserves attention in future research. Investigating the sources of individual differences in listeners' perception of L2 speech and in how accentedness is related to comprehensibility will help us gain a better understanding of the factors affecting L2 speech processing to inform L2 pronunciation instruction.

8 Conclusions

The study reported here underscores the important role of listeners' characteristics in the perception of L2 speech. Our findings indicate that non-native listeners judge speech in their own accent to be less strongly accented and to be more comprehensible than speech in an unfamiliar accent, supporting findings from previous research. In addition, we found this L1-match effect for accentedness and comprehensibility to be stronger for low- than for high-proficiency non-native listeners. Our results also indicated that the relationship between degree of accent and ease of understanding for L1-matched speech may be positive or negative as a function of listeners' L1. Altogether these findings underscore the important role of non-native listeners' native language background and L2 proficiency in the perception of non-native speech. We hope to have contributed to a better understanding of the effects of listeners' native language background and proficiency level on the perceptual assessment of accentedness and comprehensibility in L2 speech. The speech materials used are limited in that they did not allow us to investigate learners' judgements of accentedness and comprehensibility and how they relate to one another while controlling for the pronunciation-unrelated speech features that characterize L2 speech, such as speaking dysfluencies and lexical and grammatical errors. Future research should further examine L1- and proficiency-based effects on the perception of L2 speech by non-native listeners using more extemporaneous types of speech materials as well as explore further sources of individual differences in the perception of L2 speech by L2 listeners.

References

Bent, T., & Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. *The Journal of the Acoustical Society of America, 114*(3), 1600–1610. https://doi.org/10.1121/1.1603234

Crowther, D., Trofimovich, P., & Isaacs, T. (2016). Linguistic dimensions of second language accent and comprehensibility: Nonnative listeners' perspectives. *Journal of Second Language Pronunciation, 2*(2), 160–182. https://doi.org/10.1075/jslp.2.2.02cro

Crowther, D., Trofimovich, P., Isaacs, T., & Saito, K. (2015a). Does a speaking task affect second language comprehensibility? *The Modern Language Journal, 99*(1), 80–95. https://doi.org/10.1111/modl.12185

Crowther, D., Trofimovich, P., Saito, K., & Isaacs, T. (2015b). Second language comprehensibility revisited: Investigating the effects of learner background. *TESOL Quarterly, 49*(4), 814–837. https://doi.org/10.1002/tesq.203

Darcy, I. (2018). Powerful and effective pronunciation instruction: How can we achieve it? *CATESOL Journal, 30*(1), 13–45. http://www.catesoljournal.org/wp-content/uploads/2018/03/CJ30.1_darcy.pdf

Derwing, T. M., & Munro, M. J. (1997). Accent, comprehensibility and intelligibility: Evidence from four L1s. *Studies in Second Language Acquisition, 19*(1), 1–16. https://doi.org/10.1017/S0272263197001010

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. (2013). The development of L2 oral language skills in two L1 groups: A 7-year study. *Language Learning, 63*(2), 163–185. https://doi.org/10.1111/lang.12000

Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. John Benjamins. https://doi.org/10.1075/lllt.42

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. (2002). Teaching native speakers to listen to foreign-accented speech. *Journal of Multilingual and Multicultural Development, 23*(4), 245–259. https://doi.org/10.1080/01434630208666468

Eger, N. A., & Reinisch, E. (2019). The role of acoustic cues and listener proficiency in the perception of accent in nonnative sounds. *Studies in Second Language Acquisition, 41*(1), 179–200. https://doi.org/10.1017/S0272263117000377

Foote, J. A., & Trofimovich, P. (2018). Is it because of my language background? A study of language background influence on comprehensibility judgments. *Canadian Modern Language Review, 74*(2), 253–278. https://doi.org/10.3138/cmlr.2017-0011

Hayes-Harb, R., Smith, B. L., Bent, T., Bradlow, & A. R. (2008). The interlanguage speech intelligibility benefit for native speakers of Mandarin: Production and perception of English word-final voicing contrasts. *Journal of Phonetics, 36*(4), 664–679. https://doi.org/10.1016/j.wocn.2008.04.002

Isaacs, T. (2018). Shifting sands in second language pronunciation teaching and assessment research and practice. *Language Assessment Quarterly, 15*(3), 273–293. https://doi.org/10.1080/15434303.2018.1472264

Isaacs, T., & Trofimovich, P. (2011). Phonological memory, attention control, and musical ability: Effects of individual differences on rater judgments of second language speech. *Applied Psycholinguistics, 32*(1), 113–140. https://doi.org/10.1017/S0142716410000317

Isaacs, T., & Trofimovich, P. (2012). Deconstructing comprehensibility: Identifying the linguistic influences on listeners' L2 comprehensibility ratings. *Studies in Second Language Acquisition, 34*(3), 475–505. https://doi.org/10.1017/S0272263112000150

Jułkowska, I. A., & Cebrian, J. (2015). Effects of listener factors and stimulus properties on the intelligibility, comprehensibility and accentedness of L2 speech. *Journal of Second Language Pronunciation, 1*(2), 211–237. https://doi.org/10.1075/jslp.1.2.04jul

Kang, O., & Ginther, A. (2018). *Assessment in second language pronunciation.* Routledge. https://doi.org/10.4324/9781315170756

Kennedy, S., & Trofimovich, P. (2008). Intelligibility, comprehensibility, and accentedness of L2 speech: The role of listener experience and semantic context. *Canadian Modern Language Review, 64*(3), 459–489. https://doi.org/10.3138/cmlr.64.3.459

Lennon, P. (2000). The lexical element in spoken second language fluency. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 25–42). The University of Michigan Press. https://www.press.umich.edu/16106/perspectives_on_fluency

Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly, 39*(3), 369–377. https://doi.org/10.2307/3588485

Ludwig, A., & Mora, J. C. (2017). Processing time and comprehensibility judgments in non-native listeners' perception of L2 speech. *Journal of Second Language Pronunciation, 3*(2), 167–198. https://doi.org/10.1075/jslp.3.2.01lud

Mora, J. C., & Darcy, I. (2016). The relationship between cognitive control and pronunciation in a second language. In T. Isaacs, & P. Trofimovich (Eds.), *Second language pronunciation assessment: Interdisciplinary perspectives* (pp. 95–120). Multilingual Matters. https://www.multilingual-matters.com/page/detail/?k=9781783096831

Munro, M. J., & Derwing, T. M. (1995a). 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. (1995b). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech, 38*(3), 289–306. https://doi.org/10.1177/002383099503800305

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

O'Brien, M. G. (2014). L2 learners' assessments of accentedness, fluency, and comprehensibility of native and nonnative German speech. *Language Learning, 64*(4), 715–748. https://doi.org/10.1111/lang.12082

Pennycook, A. (2017). *The cultural politics of English as an international language.* Routledge. https://doi.org/10.4324/9781315225593

Pinget, A.-F., Rutger Bosker, H., 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

Saito, K., Tran, M., Suzukida, Y., Sun, H., Magne, V., & Ilkan, M. (2019). How do second language listeners perceive the comprehensibility of foreign-accented speech?: Roles of first language profiles, second language proficiency, age, experience, familiarity, and metacognition. *Studies in Second Language Acquisition, 41*(5), 1133–1149. https://doi.org/10.1017/S0272263119000226

Saito, K., Trofimovich, P., Abe, M., & In'nami, Y. (2020). Dunning-Kruger effect in second language speech learning: How does self-perception align with other perception over time? *Learning and Individual Differences*, *79*. https://doi.org/10.1016/j.lindif.2020.101849.

Saito, K., Trofimovich, P., & Isaacs, T. (2016a). 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 judgments 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

Saito, K., Webb, S., Trofimovich, P., & Isaacs, T. (2016b). Lexical correlates of comprehensibility versus accentedness in second language speech. *Bilingualism: Language and Cognition*, *19*(3), 597–609. https://doi.org/10.1017/S1366728915000255

Stibbard, R. M., & Lee, J.-I. (2006). Evidence against the mismatched interlanguage speech intelligibility benefit hypothesis. *The Journal of the Acoustical Society of America, 120*, 433–442. https://doi.org/10.1121/1.2203595

Thomson, R. I. (2018). Measurement of accentedness, intelligibility and comprehensibility. In K. Okim & A. Ginther (Eds.), *Assessment in second language pronunciation* (pp. 11–29). Routledge. https://doi.org/10.4324/9781315170756-2/measurement-accentedness-intelligibility-comprehensibility-ron-thomson

Trofimovich, P., & Baker, W. (2006). Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition, 28*(1), 1–30. https://doi.org/10.1017/S0272263106060013

Winke, P., Gass, S., & Myford, C. (2013). 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

Joan C. Mora is Associate Professor in the Department of Modern Languages and Literatures and English Studies at the University of Barcelona (UB), Spain. His research has examined the acquisition of L2 phonology and the role of contextual and individual factors in the development of L2 speech and oral fluency.

Perception and Recoverability of Modified English L2 Codas

Ali S. Alelaiwi and Steven H. Weinberger

Abstract Previous research has shown that when L2 learners are faced with illegal structures, they employ various modification strategies to avoid such structures. This chapter reports on a study that examined the perception of two of these strategies: deletion and epenthesis. The participants were presented with monosyllabic words with codas modified by either deletion or epenthesis and asked if they favored one modification strategy over the other. A hundred and thirty-seven listeners from three different language backgrounds—English, Spanish, and Japanese—were recruited to complete this perceptual task. Our findings revealed that epenthesis was significantly preferred over deletion regardless of the listeners' L1, which provides support for the Recoverability Principle.

Keywords Recoverability · Syllable modification · Epenthesis · Coda perception · Sonority

1 Introduction

Previous studies have shown that when second language (L2) learners are faced with structures that are illegal in their first language (L1), they tend to simplify such structures (Abrahamsson, 2003; Hansen, 2004; Osburne, 1996; Sato, 1984; Weinberger, 1994; Yavaş, 2011). This chapter examines two different strategies of syllable structure simplification, namely, consonant deletion and vowel epenthesis, from a perceptual perspective. Specifically, it reports on a perception study that we conducted to investigate the *Recoverability Principle* (Weinberger, 1994), which suggests that epenthesis is functionally superior to deletion because it results in relatively less ambiguous structures.

A. S. Alelaiwi (✉) · S. H. Weinberger
Department of English, George Mason University, Fairfax, VA, USA
e-mail: asalelaiwi@nu.edu.sa

S. H. Weinberger
e-mail: weinberg@gmu.edu

V. G. Sardegna and A. Jarosz (eds.), *Theoretical and Practical Developments in English Speech Assessment, Research, and Training*, Second Language Learning and Teaching,
https://doi.org/10.1007/978-3-030-98218-8_8

This chapter is organized as follows. Section 2 reviews related studies on the production of English codas and provides the necessary theoretical background. Section 3 describes the purpose of our study, and the languages and predictions we tested. Section 4 explains the methodology we used to collect and analyze the perceptual data, including information about the participants, stimuli, and procedures. Section 5 presents and discusses our findings. Finally, Sect. 6 offers our conclusions and identifies areas of future research.

2 Literature Review

2.1 Theoretical Framework

2.1.1 The Recoverability Principle

Weinberger (1987) proposed the Recoverability Principle to guide preference structures resulting from phonological operations such as deletion and epenthesis. For example, if we examine a word with a CVC syllable structure such as *lead*, there are two possible simplification outcomes for the target word in (1):

(1) Target word	Deleted form	Epenthesized form
lead [lid]	[li]	[lidə]

The deleted form results in more ambiguity since it could be interpreted as *Lee* (proper name)*, leaf, leave, lean, lead, leak, leash, lease*, etc. The epenthesized form [lidə], on the other hand, results in less potential ambiguity because it can only be interpreted as *lead,* or *leader* if the person speaks a variety of English where the deletion of final [ɹ] is acceptable. Hence, as this example illustrates, epenthesis is preferred when it comes to meaning preservation (Weinberger, 1994). The Recoverability Principle is formally expressed in (2):

(2) Modifications resulting in recoverable outputs are preferred over modifications resulting in nonrecoverable outputs. (Weinberger, 1987)

According to Weinberger (1994), the Recoverability Principle is part of a universal grammar that matures following a preset schedule. He argues that children do not employ epenthesis as a simplification strategy because the Recoverability Principle is not yet active due to children's limited lexicon. By the time it becomes active, children whose native language allows coda consonants are already capable of producing the complex structures. Based on this claim, it can be predicted that adult L2 learners will employ epenthesis as their predominant simplification strategy since the Recoverability Principle is presumably active. However, studies investigating cluster simplification strategies show that this is not always the case (e.g., Abrahamsson, 2003; Benson, 1988; Sato, 1984; Weinberger, 1987).

2.1.2 Sonority

It is fairly well established that, cross-linguistically, the segments within a syllable pattern in a certain manner based upon sonority (Broselow & Finer, 1991; Carlisle, 2001; Clements, 1990; Hansen, 2001; Parker, 2002; Tropf, 1986). A universally preferred syllable is one in which the nucleus is the most sonorous constituent, whereas the sonority of the other segments in the syllable (coda and onset segments) decreases continuously outward from the nucleus. This organization of segments within a syllable is referred to as the *Sonority Sequencing Principle* (Clements, 1990; Parker, 2002), which is formally expressed in (3):

(3) Between any member *x* of a syllable and the syllable peak *p*, only sounds of higher sonority rank than *x* are permitted. (Kar, 2010)

One-member onsets and codas by definition must adhere to the Sonority Sequencing Principle since they must be comprised of segments that are less sonorant than the nucleus (Carlisle, 2001). However, one-member onsets and codas differ dramatically from each other according to which segments are preferred. If an onset consists of one segment, there is a universal tendency for this segment to be low in sonority, which results in obstruents being preferred over sonorants in that position. The reverse is true for codas where one-member codas that are high in sonority are preferred. This generalization is referred to as the *Sonority Dispersion Principle*, which requires sonority to be maximally dispersed in the initial demisyllable and minimally dispersed in the final demisyllable (Clements, 1990). In other words, the sharper the rise in sonority from the beginning of the syllable to the nucleus, the better the syllable. The opposite is true for the end of the syllable in which the fall of sonority from the nucleus needs to be minimal.

A number of different sonority scales has been proposed in the literature, but in this chapter, the scale in (4) will be used as a starting point. Each of the segments forming the syllable will take its place on this scale, according to its sonority properties.

(4) Stops < Fricatives < Nasals < Liquids < Glides < Vowels. (Morelli, 2003)

Affricates are not usually included in most common scales of sonority due to their complex nature. Some researchers suggest that they have the same sonority profile as stops (Bolinger, 1962; Cardona, 1988). Others suggest that affricates are between stops and fricatives, as in (5).

(5) Stops < Affricates < Fricatives. (Goldsmith, 1995; Katamba, 1989; Puppel & Fisiak, 1992)

In our study (see Sect. 3), we treated affricates as a separate sonority group due to their debatable classification. Thus, we followed the scale in (6) and invoked sonority as one of the contributing factors in the deletion/epenthesis asymmetry.

(6) Stops < Affricates < Fricatives < Nasals < Liquids < Glides < Vowels.

2.2 Previous Studies on the Production of English Codas

Several studies have investigated the production of English codas by L2 learners. However, all have focused on production (Benson, 1988; Hansen, 2004; Sato, 1984; Wang, 1995; Weinberger, 1987; Yavaş, 2011). For example, Sato (1984) conducted a longitudinal study examining the production of two-member codas in spontaneous speech samples of two Vietnamese children at three different time points during a period of 10 months. The results showed that, of all non-target forms, 2.43% were modified by feature change, 84.14% were modified by cluster reduction, and 13.41% were modified by cluster deletion. Sato concluded that native speakers of Vietnamese tend to modify codas by single consonant deletion rather than epenthesis or deletion of the entire cluster.

Similarly, Benson (1988) examined the production of monosyllabic English words in informal conversations by two Vietnamese speakers of English. The results showed that the first participant had 396 attempted productions of monosyllabic closed syllable target words (CVC), of which 67 cases (16.91%) were modified by deletion. The second participant had 141 attempted productions of monosyllabic closed syllable target words (CVC), of which 25 cases (17.7%) were modified by deletion. Similar to Sato's study, Benson (1988) pointed out that none of the two participants used epenthesis as a modification strategy.

Weinberger (1987) examined the production of one-member, two-member, and three-member word-final codas by four intermediate Mandarin speakers of English and found that they exhibited 50% epenthesis and 50% deletion. Weinberger suggested that this finding may have been due to the participants' overall English proficiency. He argued that adult L2 learners with a more developed knowledge of the target lexicon could be more sensitive to the Recoverability Principle. Indeed, this has been shown to occur developmentally in L2 acquisition by Abrahamsson (2003).

Yavaş (2011) investigated the acquisition of two-member English coda clusters by 19 native speakers of Spanish. He looked at the production of 24 monosyllabic and mono-morphemic English words and concluded that L1 Spanish speakers only modified these target words by deletion. There were 139 cases of deletion out of 456 possible cases (30% deletion). This finding actually represents a challenge for the generalizability of Weinberger's (1987) proposal. If we apply Weinberger's proposal to an English word like *milk* [mɪlk] (i.e. a word with a two-member coda cluster similar to those used in Yavaş's study), we should end up with [mɪl.kV] or [mɪ.lV.kV] (i.e., the epenthesis form). Nevertheless, such outcomes were not produced by the Spanish speakers. However, it is worth pointing out that Yavaş (2011) did not mention anything regarding the learners' overall English proficiency. Thus, it is possible that they were non-advanced learners of English. Consequently, it is also possible that they had not yet developed the adequate linguistic knowledge of the target lexicon that would lead them to employ the Recoverability Principle proposed by Weinberger (1987). Nevertheless, the possibility that Spanish learners have a general preference for coda deletion as a simplification strategy cannot be ignored, either. That is to say, regardless of the number of syllables in the target words or their overall English proficiency level, it is possible to predict, based on Yavaş's findings, that Spanish learners of English will choose deletion as the main strategy when faced with illegal codas.

The tendency of a certain language to systematically apply one repair strategy over another is not entirely unusual. For example, it has been observed that the English interdentals [θ, ð] are replaced either by [t] or [s], depending on the speaker's L1. Generally, the segment used for substitution is consistent for speakers of a given language. For example, L1 German speakers are reported to use [s], whereas L1 Russian speakers use [t] systematically (Lombardi, 2003). This area of research and, specifically, how repair strategies differ across languages and between speakers and listeners, needs further investigation. Based on the observation that speakers of a certain language may systematically apply one repair strategy over another and that it is not yet known whether speakers and listeners of the same language would use the same repair strategy, we conducted a study that examined the perception of structures modified by either epenthesis or deletion by listeners of three different L1s.

3 The Study

This study examined the preference between two strategies (vowel epenthesis and consonant deletion) of syllable structure simplification using a perceptual task (Boudaoud & Cardoso, 2009; Eckman, 1991; Edge, 1991). As far as we know, at the time of this study no previous study had examined the difference between epenthesis and deletion using a perceptual task. By conducting a perception experiment, we addressed this gap in the literature.

Based on functional approaches to phonology and phonetics, speakers are governed by two conflicting forces: (a) their tendency to minimize articulatory effort, and (b) their need to maximize intelligibility (Abrahamsson, 2003). The first is based on the speaker, and it manifests itself in phonological processes that result in unmarked structures. The second is oriented towards the needs of the listener, and it manifests itself in the need to maintain distinctness and understandability. With respect to the processes under examination (deletion and epenthesis), if adult speakers are to minimize articulatory complexity, deletion should be the strategy of choice. If, however, adult speakers ultimately want to be understood, they should employ epenthesis rather than deletion since it accommodates the listeners' needs by maintaining relevant information and avoiding ambiguous forms, as predicted by the Recoverability Principle (Weinberger, 1994). Nevertheless, as evidenced from the previously mentioned production studies, epenthesis is not always the strategy of choice by adult speakers. Thus, conducting a perception experiment allowed us test the implications of the Recoverability Principle on listeners by factoring out the vagaries of articulation concerns often associated with production experiments.

3.1 Languages Under Examination

In our study, we examined the perception of structures modified by either epenthesis or deletion by listeners whose L1 was Japanese, Spanish, and English. Prior research has shown that Japanese listeners perceive a vowel when presented with words containing illegal structures even when the vowel was not actually present (Dupoux et al., 1999).

Furthermore, Japanese is more restrictive in the range of coda consonants it allows compared to English. For example, Japanese only allows codas in two cases: (a) if the segment is a nasal, or (b) if it is the first part of a geminate which can only appear word-medially (Tsuchida, 1995). In contrast, as previously discussed, Yavaş (2011) has shown that Spanish speakers favor deletion when it comes to modifying illegal codas with two consonants. Since Yavaş (2011) only examined two-member coda clusters in a production study, it is unknown whether this finding can be generalized to the perception of singleton codas. Our study attempted to fill this gap. Similar to Japanese, Spanish has a very limited set of coda possibilities. It only allows five consonants in the coda position [d, s, n, ɾ, l] (Núñez-Cedeño et al., 2014). English, on the other hand, allows for a much larger set. Most English consonants can occur in the coda position [p, b, t, d, k, ɡ, m, n, ŋ, f, v, θ, ð, s, z, ʃ, ɹ, l, dʒ, tʃ] (Gregová, 2010). Finally, since the stimuli in our study consisted of English words modified by either epenthesis or deletion, the perception of English listeners was examined as a control group.

3.2 Predictions

Based on prior research, we hypothesized that:

- The Recoverability Principle operates in the perception grammar.
  - Words modified by epenthesis will be chosen more frequently by adult listeners of all languages (Weinberger, 1994).
- Sonority of the coda consonant will influence the modification strategy.
  - If the original word ends on a segment with low sonority (e.g., [t]), listeners will choose the word modified by epenthesis. This is because epenthesis creates another syllable in which the segment previously in the coda will be the onset of the new syllable, and onsets with low sonority are preferred (Clements, 1990).
  - If the original word ends on a segment with high sonority, listeners will choose the word modified by deletion.
- There will be native language bias.
  - Spanish listeners will choose words modified by deletion more often (Yavaş, 2011).

- Japanese listeners will choose words modified by epenthesis more often (Dupoux et al., 1999).

- Proficiency matters.
  - Listeners with higher English proficiency will choose words modified by epenthesis more frequently (Weinberger, 1994).

4 Methodology

4.1 Participants

Our study examined listeners from three different language backgrounds: English, Japanese and Spanish. A total of 137 listeners were recruited via Amazon Mechanical Turk, and each was given \$1.50 as compensation. Participants who reported having hearing or speaking issues were excluded from the study. In addition, Japanese and Spanish participants were asked to self-rate their English proficiency and frequency of English use using a five-point scale (1 = very low proficiency/frequency of use, and 5 = high proficiency/frequent language use). The percentages were calculated by summing up all the proficiency scores for each language group and then dividing the actual outcome by the total possible proficiency score for that particular language. The obtained decimal value was then multiplied by 100 to get the percentage. Table 1 displays participants' demographic information and obtained scores.

The three groups were similar in terms of listeners' mean age. In addition, independent-samples t-tests revealed that there were no significant differences with respect to age of onset, $t(74) = 0.648$, $p = 0.519$, and self-reported English proficiency, $t(74) = 0.346$, $p = 0.731$, between the Japanese and Spanish groups. However, the Japanese participants reported a higher frequency of English use (70.52%) compared to the Spanish participants (48.75%), and this difference was statistically significant, $t(37) = –5.29$, $p < 0.001$. Such difference could be related to participants' differences in length of residence, with the Japanese participants averaging 11 years and the Spanish averaging only two years, which was also found statistically significant: $t(74) = 5.69$, $p < 0.001$.

Table 1 Participants' demographic information

L1	N	Age	Gender	Age of onset	Length of residency	English proficiency	Frequency of English use
English	51	(21—70) $M = 29.47$	M = 27 F = 24	NA	NA	NA	NA
Japanese	38	(18—43) $M = 31.83$	M = 21 F = 17	(3—21) $M = 8.39$ $SD = 4.99$	(0—35) $M = 11.1$ $SD = 8.25$	84.73% $M = 4.24$ $SD = 0.542$	70.52% $M = 3.53$, $SD = 0.830$
Spanish	48	(19—52) $M = 31.58$	M = 34 F = 14	(1—25) $M = 7.71$ $SD = 4.18$	(0—25) $M = 2.16$ $SD = 4.69$	82.5% $M = 4.18$ $SD = 0.766$	48.75% $M = 2.50$ $SD = 0.191$

4.2 Stimuli

The stimuli consisted of 38 monosyllabic monomorphemic English nouns with a CVC syllable structure. In each session, the participants were presented with two modified forms of each word of the original 38 words, one with vowel epenthesis (CVCV) and the other with consonant deletion (CV). This means that they listened to 76 (38 $\times$ 2) forms of the experimental words, and they had to choose one variant per question.

The experimental words were chosen to cover all consonants that can occur in the English coda position. Nineteen different coda consonants were included: [p, b, t, d, k, g, f, v, θ, s, z, ʃ, tʃ, dʒ, m, n, ŋ, l, ɹ]. One consonant, the voiced interdental fricative [ð], was not included because it was not found in coda positions in monosyllabic nouns. Each of the coda consonants appeared twice in two different words. This resulted in a total of 38 target words per participant. In addition, the participants were presented with 15 nonexperimental words (fillers). These fillers consisted of words with onset clusters, such as *flake*, for which each question contained two forms: the original form [fleɪk] and another that was modified by deleting one member of the onset cluster [leɪk]. Of these 15 fillers, three were used in a brief training session.

All experimental words and fillers (see Appendix) were produced by a phonetically-trained male native speaker of English. The speaker's age was 62. He was born in Pittsburgh, PA, USA, and he reported having knowledge of Mandarin. The speaker was asked to produce two forms of each word. For the words modified by deletion, he was asked to drop the coda. For the words modified by epenthesis, he was asked to add the vowel [ə]. The words were recorded with a 44.1 kHz sampling rate using Zoom H2 Handy Recorder in the Acoustics Lab at George Mason University. The recorded words were normalized at 3db, and the epenthesized vowels were checked for duration consistency.

4.3 Procedure

The experiment was designed in Qualtrics, and then it was linked to Amazon Mechanical Turk. All participants first completed a consent form. Once they agreed, they were asked to provide some demographic information: native language, age, gender, English proficiency, frequency of English use, length of residency, age of onset, place of birth and method of learning English (naturally or academically). Participants who did not meet the requirements for the study, such as those who reported having hearing problems, were excluded from the analysis. All participants were required to wear headsets and enter the model name of the headset they were using. Those who failed to provide this information were excluded from the study. Once they completed the background information, they were presented with three stimuli containing filler words as part of a training session. After the familiarization trials, the actual experiment started. Each participant was presented with 50 stimuli containing

38 experimental words and 12 fillers with corresponding pictures in a randomized order (see Appendix). For each word, participants heard two forms. For example, for the English word *couch*, participants were shown a picture of a couch, and heard the two modified forms [kaʊʧə] and [kaʊ] denoting the picture. They were instructed to choose the word that best matched the picture based on their judgment.

4.4 Analysis

Jamovi (Datalab.cc, n.d.) was used to perform the statistical analysis. A mixed model regression test was conducted to see if the listeners' native language and the sonority of coda consonants significantly influenced the choice of repair strategy (deletion vs. epenthesis). In this model, *deletion* was set as the dependent variable; *language*, *sonority* and *sonority*language* (the interaction between language and sonority) were the fixed factors; *participant* and *word* were the random structures. We also performed a post-hoc test to compare sonority profiles. We report the results in Sect. 5.

5 Results and Discussion

The results of the mixed model regression test indicate that the choice of strategy (epenthesis vs. deletion) was significantly influenced by the participants' native language [$F(2, 142) = 14.12$, $p < 0.001$] and the sonority profile [$F(4, 33) = 2.86$, $p = 0.038$]. The interaction between language and sonority was also statistically significant [$F(8, 5024) = 4.88$, $p < 0.001$].

In order to examine the combined performance of all groups in relation to specific sonority profiles, a Bonferroni post-hoc analysis was conducted. The results showed that liquid was the only sonority level that significantly exhibited deletions ($p = 0.005$) compared to other sonority levels. Furthermore, when we examine the performance of a specific language in relation to specific sonority profiles compared to the other languages, we find that Spanish participants had a significant tendency for deleting stops ($p < 0.001$), fricatives ($p = 0.017$), nasals ($p = 0.008$), and liquids ($p = 0.018$). Moreover, Japanese participants had a significant tendency for deleting stops ($p = 0.017$) and nasals ($p = 0.027$) (see Table 2).

Figure 1 displays the deletion frequency for all examined languages across all sonority profiles.

In the following subsections, we discuss the results of each language group in more detail.

Table 2 Fixed effects parameter estimates

Names	Effect	Estimate	95% CI			*Df*	*T*	*p*
			SE	Lower	Upper			
(Intercept)	(Intercept)	0.09186	0.01405	0.06433	0.11939	98.8	6.540	< 0.001
Language1	Japanese—(English, Japanese, Spanish)	0.03884	0.01554	0.00839	0.06929	141.6	2.500	0.014
Language2	Spanish—(English, Japanese, Spanish)	0.03766	0.01463	0.00898	0.06634	141.6	2.573	0.011
Sonority1	fricative—(affricate, fricative, liquid, nasal, stop)	– 0.02292	0.01589	– 0.05406	0.00822	33.1	– 1.443	0.158
Sonority2	liquid—(affricate, fricative, liquid, nasal, stop)	0.07049	0.02356	0.02430	0.11667	33.1	2.991	0.005
Sonority3	nasal—(affricate, fricative, liquid, nasal, stop)	– 0.01505	0.02010	– 0.05444	0.02434	33.1	– 0.749	0.459
Sonority4	stop—(affricate, fricative, liquid, nasal, stop)	0.00872	0.01589	– 0.02242	0.03986	33.1	0.549	0.587
Language1 ✻ Sonority1	Japanese—(English, Japanese, Spanish) ✻ fricative—(affricate, fricative, liquid, nasal, stop)	0.00406	0.00881	– 0.01321	0.02133	5024.0	0.461	0.645
Language2 ✻ Sonority1	Spanish—(English, Japanese, Spanish) ✻ fricative—(affricate, fricative, liquid, nasal, stop)	– 0.01979	0.00830	– 0.03606	– 0.00352	5024.0	– 2.384	0.017
Language1 ✻ Sonority2	Japanese—(English, Japanese, Spanish) ✻ liquid—(affricate, fricative, liquid, nasal, stop)	0.00276	0.01307	– 0.02286	0.02838	5024.0	0.211	0.833
Language2 ✻ Sonority2	Spanish—(English, Japanese, Spanish) ✻ liquid—(affricate, fricative, liquid, nasal, stop)	0.02917	0.01231	0.00503	0.05330	5024.0	2.369	0.018

(continued)

Table 2 (continued)

Names	Effect	Estimate	95% CI			*Df*	*T*	*p*
			SE	Lower	Upper			
Language1 ✻ Sonority3	Japanese—(English, Japanese, Spanish) ✻ nasal—(affricate, fricative, liquid, nasal, stop)	0.02470	0.01115	0.00285	0.04655	5024.0	2.216	0.027
Language2 ✻ Sonority3	Spanish—(English, Japanese, Spanish) ✻ nasal—(affricate, fricative, liquid, nasal, stop)	– 0.02766	0.01050	– 0.04824	– 0.00708	5024.0	– 2.634	0.008
Language1 ✻ Sonority4	Japanese—(English, Japanese, Spanish) ✻ stop—(affricate, fricative, liquid, nasal, stop)	– 0.02100	0.00881	– 0.03827	– 0.00373	5024.0	– 2.383	0.017
Language2 ✻ Sonority4	Spanish—(English, Japanese, Spanish) ✻ stop—(affricate, fricative, liquid, nasal, stop)	0.03364	0.00830	0.01737	0.04991	5024.0	4.053	< 0.001

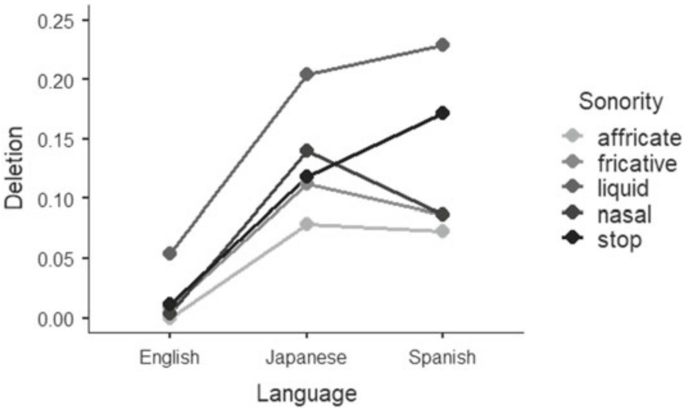

Fig. 1 Deletion-choices based on language and sonority

5.1 English Participants

A total of 51 native speakers of English participated in this study. They were each presented with 38 experimental words and were asked to choose between words that were modified by epenthesis or deletion. This resulted in a total of 1,938 tokens. Out of 1,938 tokens, words modified by epenthesis were chosen 1,914 times (98.76%) and words modified by deletion were only chosen 24 times (1.23%).

The difference between epenthesis and deletion was statistically significant ($\beta = 0.091$, $p < 0.001$). Native speakers of English had a dominant preference for words modified by epenthesis, which provides support for the Recoverability Principle that states that words modified by epenthesis are easier to disambiguate. Based on these findings, we could argue that in real communication, with all else being equal and from a listener's perspective, native English speakers would find words modified by epenthesis preferable to those modified by deletion. And as we continue to argue in this chapter, language users pay attention to lexical ambiguity.

We also examined the specific segments in the coda position. Based on a sonority perspective, we predicted that, if the coda had a segment with low sonority, listeners would choose the word modified by epenthesis since the original coda would be transformed into an onset where segments with lower sonority are preferred. On the contrary, if the original word ended on a segment with high sonority, listeners would choose the word modified by deletion based on the Sonority Dispersion Principle (Clements, 1990). Table 3 lists the exact number of epenthesis and deletions produced by the English participants based on sonority profile. The column labeled *possible total* indicates the total number of tokens for each sonority profile. This number is the result of the original number of words ending in segments in a particular sonority profile multiplied by the number of participants.

As mentioned previously and shown in Table 3, epenthesis was predominantly more frequent than deletion for the English participants. When it comes to sonority,

Table 3 English: Total epenthesis vs. deletion based on sonority profile

Sonority	Possible total	Epenthesis	Deletion
Stops	$12 \times 51 = 612$	605 (98.85%)	7 (1.14%)
Affricates	$4 \times 51 = 204$	204 (100%)	0 (0%)
Fricatives	$12 \times 51 = 612$	607 (99.18%)	5 (0.81%)
Nasals	$6 \times 51 = 306$	305 (99.67%)	1 (0.32%)
Liquids	$4 \times 51 = 204$	193 (94.60%)	11 (5.69%)
Totals	1938 (100%)	1914 (98.76%)	24 (1.23%)

Table 4 Post hoc comparisons—English ✻ Sonority

Language	Sonority	Language	Sonority	Difference	*SE*	Test	*df*	Pbonferroni
English	Liquid	English	Nasal	0.05065	0.0394	1.2848	Inf	1.000
English	Liquid	English	Stop	0.04248	0.0353	1.2047	Inf	1.000
English	Nasal	English	Stop	–0.00817	0.0305	–0.2675	Inf	1.000
English	Affricate	English	Liquid	–0.05392	0.0432	–1.2485	Inf	1.000
English	Affricate	English	Nasal	–0.00327	0.0394	–0.0829	Inf	1.000
English	Affricate	English	Stop	–0.01144	0.0353	–0.3244	Inf	1.000
English	Affricate	English	Fricative	–0.00817	0.0353	–0.2317	Inf	1.000
English	Fricative	English	Liquid	–0.04575	0.0353	–1.2974	Inf	1.000
English	Fricative	English	Nasal	0.00490	0.0305	0.1605	Inf	1.000
English	Fricative	English	Stop	–0.00327	0.0249	–0.1311	Inf	1.000

liquids exhibited the greatest number of deletions totaling 5.69%. We further examined the 11 cases of liquid deletions and found that out of the 11 cases, 10 cases (90.9%) were instances of [ɹ] deletions. This higher percentage of [ɹ] deletions compared to other consonants could be attributed to the acceptability of [ɹ] deletions in many dialects of English.

To see if the differences between the sonority profiles were statistically significant, a post-hoc test was performed. The results showed that the English participants generally preferred epenthesis regardless of the sonority profile. Table 4 also shows that the slightly higher percentage of liquid deletions is not statistically significant. Based on these findings, we can conclude that native English speakers find words modified by epenthesis preferable regardless of sonority.

5.2 Japanese Participants

There was a total of 38 Japanese participants in this study. Each participant was presented with 38 experimental words and were asked to choose between words that were modified by epenthesis or deletion. This resulted in a total of 1,444 tokens.

Out of these 1,444 tokens, words modified by epenthesis were chosen 1,264 times (87.53%) and words modified by deletion were chosen 180 times (12.46%). This difference between epenthesis and deletion was statistically significant ($\beta = 0.038$, $p = 0.014$), which provides additional support for the Recoverability Principle (i.e., words modified by epenthesis are preferred).

We also examined the specific segments that underwent deletion. As Table 5 shows, similar to the English sample, epenthesis was predominantly more frequent across all sonority types.

With respect to deletions, the most sonorous categories, namely liquids and nasals, had slightly higher deletions compared to fricatives and stops. Specifically, liquids exhibited the highest percentage (20.39%), and nasals came immediately after (14.03%). A post-hoc test was conducted to see if there was a significant interaction between deletions and sonority. Table 6 shows that the Japanese participants' rate of deletion was not significantly influenced by the sonority profile. This indicates that, similar to English, Japanese speakers find words modified by epenthesis preferable regardless of sonority.

Compared to the Japanese sample, the English sample had a relatively greater preference for epenthesis. The English participants chose epenthesis 98.76% of the time compared to 87.53% in the Japanese sample. A post-hoc test showed that this

Table 5 Japanese: Total epenthesis vs. deletion based on sonority profile

Sonority	Possible Total	Epenthesis	Deletion
Stops	$12 \times 38 = 456$	402 (88.15%)	54 (11.84%)
Affricates	$4 \times 38 = 152$	140 (92.10%)	12 (7.89%)
Fricatives	$12 \times 38 = 456$	405 (88.81%)	51 (11.18%)
Nasals	$6 \times 38 = 228$	196 (85.96%)	32 (14.03%)
Liquids	$4 \times 38 = 152$	121 (79.60%)	31 (20.39%)
Totals	1444 (100%)	1264 (87.53%)	180 (12.46%)

Table 6 Post hoc comparisons—Japanese ✻ Sonority

Language	Sonority	Language	Sonority	Difference	*SE*	Test	*df*	pbonferroni
Japanese	Liquid	Japanese	Nasal	0.06360	0.0414	1.5355	Inf	1.000
Japanese	Liquid	Japanese	Stop	0.08553	0.0370	2.3088	Inf	1.000
Japanese	Nasal	Japanese	Stop	0.02193	0.0321	0.6836	Inf	1.000
Japanese	Affricate	Japanese	Liquid	–0.12500	0.0454	–2.7552	Inf	0.616
Japanese	Affricate	Japanese	Nasal	–0.06140	0.0414	–1.4826	Inf	1.000
Japanese	Affricate	Japanese	Stop	–0.03947	0.0370	–1.0656	Inf	1.000
Japanese	Affricate	Japanese	Fricative	–0.03289	0.0370	–0.8880	Inf	1.000
Japanese	Fricative	Japanese	Liquid	–0.09211	0.0370	–2.4864	Inf	1.000
Japanese	Fricative	Japanese	Nasal	–0.02851	0.0321	–0.8886	Inf	1.000
Japanese	Fricative	Japanese	Stop	–0.00658	0.0262	–0.2512	Inf	1.000

difference was statistically significant ($\beta = 0.091$, $p < 0.001$). This difference could be attributed to the fact that the non-native Japanese participants did not possess the same linguistic proficiency (they self-reported an average of 84.73%) as the English participants, which may have prevented them from employing the Recoverability Principle as frequently as the English participants.

5.3 Spanish Participants

There was a total of 48 participants in this study. Each participant was presented with 38 experimental words and had to choose between words that were modified by epenthesis or deletion. This resulted in a total of 1,824 tokens. Out of these 1,824 tokens, words modified by epenthesis were chosen 1,593 times (87.33%) and words modified by deletion were chosen 231 times (12.66%). The difference between epenthesis and deletion was statistically significant ($\beta = 0.037$, $p = 0.011$). This finding provides additional support for the Recoverability Principle as it indicates that Spanish speakers found English words modified by epenthesis easier to disambiguate.

We also examined the specific segments that had undergone deletion (see Table 7).

Epenthesis was the most frequent regardless of the sonority profile, reaching a total of 1593 cases (87.33%). This is consistent with what we found in the English and Japanese samples. Also, similar to the other groups, liquids exhibited the highest percentage of deletions (22.91%). A post-hoc test revealed that the Spanish participants deleted liquids significantly more than nasals ($p = 0.036$), fricatives ($p = 0.007$) and affricates ($p = 0.036$). However, there was no significant difference between liquids and stops. A possible explanation for the high frequency of liquid deletions could be due to the acceptability of [ɹ] deletions in many English dialects. A close inspection at the types of deletions uncovered that out of 44 deletions, 38 cases (86.36%) were instances of [ɹ] deletions whereas only 6 cases (13.63%) were instances of [l] deletions. It is worth pointing out that, unlike the Japanese participants, the nasal was not the coda type with the second most deletions in the Spanish group. In Spanish, stop deletions amounted to 17.36%, which comes immediately after liquids (22.91%).

Table 7 Spanish: Total epenthesis vs. deletion based on sonority profile

Sonority	Possible Total	Epenthesis	Deletion
Stops	$12 \times 48 = 576$	476 (82.63%)	100 (17.36%)
Affricates	$4 \times 48 = 192$	178 (92.70%)	14 (7.29%)
Fricatives	$12 \times 48 = 576$	528 (91.66%)	48 (8.33%)
Nasals	$6 \times 48 = 288$	263 (91.31%)	25 (8.68%)
Liquids	$4 \times 48 = 192$	148 (77.08%)	44 (22.91%)
Totals	1824 (100%)	1593 (87.33%)	231 (12.66%)

Table 8 Post hoc comparisons—Spanish ✻ Sonority

Language	Sonority	Language	Sonority	Difference	*SE*	Test	*df*	pbonferroni
Spanish	Liquid	Spanish	Nasal	0.14236	0.0398	3.5772	Inf	0.036
Spanish	Liquid	Spanish	Stop	0.05729	0.0356	1.6095	Inf	1.000
Spanish	Nasal	Spanish	Stop	–0.08507	0.0308	–2.7596	Inf	0.608
Spanish	Affricate	Spanish	Liquid	–0.15625	0.0436	–3.5841	Inf	0.036
Spanish	Affricate	Spanish	Nasal	–0.01389	0.0398	–0.3490	Inf	1.000
Spanish	Affricate	Spanish	Stop	–0.09896	0.0356	–2.7801	Inf	0.571
Spanish	Affricate	Spanish	Fricative	–0.01389	0.0356	–0.3902	Inf	1.000
Spanish	Fricative	Spanish	Liquid	–0.14236	0.0356	–3.9994	Inf	0.007
Spanish	Fricative	Spanish	Nasal	–1.60e–15	0.0308	–5.19e–14	Inf	1.000
Spanish	Fricative	Spanish	Stop	–0.08507	0.0252	–3.3798	Inf	0.076

Table 8 shows that liquid deletions were not statistically significant when compared to other sonority profiles. Liquid deletions approached significance, nevertheless, only when compared to fricatives ($p = 0.007$). This finding is interesting because, based on sonority, one would not expect stops to be the second highest to exhibit deletions since they make ideal onsets. A possible explanation for this outcome could be that stops are the least marked segments (de Lacy, 2002). Unmarked segments may be easier to produce due to their articulatory simplicity, yet they have less perceptual salience. Such features make unmarked segments subject to change. Hume (2004) points out that phonologically unmarked segments in a system are considered to be the least stable phonetically. That is, they are most likely to undergo processes such as reduction, deletion, and assimilation. Our findings seem to support this explanation.

5.4 *Findings Across Groups*

We also looked at the results of epenthesis frequency across the three language groups. We found that, similar to the Japanese listeners, Spanish participants had a relatively lower epenthesis frequency (87.33%) compared to the English sample (98.76%). A post hoc-test indicated that the difference between the Spanish and English samples was statistically significant ($p < 0.001$). Just as with the Japanese listeners, this difference could be attributed to the difference in proficiency levels since Spanish participants reported an average English proficiency of 82.5%. Thus, the findings from the two L2 speaker groups (the Spanish and Japanese participants) seem to suggest that English proficiency influences the choice of modification strategy. These findings are consistent with previous production studies (Abrahamsson, 2003; Weinberger, 1987).

Moreover, another interesting finding is that there was no statistical significance between the Spanish and Japanese groups with respect to their choice of forms modified by epenthesis: Spanish speakers chose these forms 87.33% of the time, whereas Japanese speakers chose them 87.53% of the time. Based on Yavaş (2011), we predicted that Spanish participants would choose deletion more frequently than epenthesis as the Spanish participants in his study predominantly deleted consonants to modify two-member coda clusters in their production of English words. Yet, this was not the case in our perceptual study. A possible explanation is that the choice of strategy is dependent upon the length of the coda such that singleton codas are modified by epenthesis whereas two-member codas are modified by deletion. If that is the case, then the findings in Yavaş (2011) with two-member codas cannot be extended to the singleton codas used in our study. Another possible explanation is that, since both the Japanese and Spanish listeners in our study reported a similar high English proficiency, they may have reached the same level of competence that is needed to exploit the Recoverability Principle. In contrast, the participants in Yavaş's study may not have had the English level needed to exploit the Recoverability Principle. This hypothesis needs to be tested with further research as, unfortunately, Yavaş did not report the English proficiency level of his study participants. These are all issues for future study.

5.5 *Implications*

The significance of this research is that it contributed to the body of knowledge in linguistics by examining the Recoverability Principle using a perceptual perspective. Previous research on the Recoverability Principle was only done on production data. We believe that this type of perceptual study gives us further insight into the grammatical knowledge that L2 learners have about the lexicon of their target language without the complication of articulatory behavior.

Furthermore, we have shown that the two modification strategies of epenthesis and deletion are not equal when examined from the perspective of listeners—that is, the choice of behavior is dependent upon other factors. This finding may serve as a useful diagnostic for language teachers and language assessment professionals. For example, it is true that language learners proceed through the stages of syllable-simplification behaviours in their production; that is, first using consonant deletion, and then using vowel epenthesis. Yet, we have found that this is not necessarily the case in perception. We therefore believe that it is beneficial for a teacher/assessor to understand that their students' production does not faithfully reflect their true grammatical (perceptual) knowledge of the target lexicon.

6 Conclusion

The purpose of this study was to examine the preference between two common modification strategies, vowel epenthesis and consonant deletion, using a perception experiment. Specifically, we hypothesized that if the Recoverability Principle plays a role in determining the modification strategy employed, epenthesis will be significantly more preferred by listeners compared to deletion. To test this hypothesis, we targeted participants from three different linguistic backgrounds: English, Japanese, and Spanish. The results showed that epenthesis was significantly more preferred in all examined languages.

We also wanted to test the hypothesis that sonority would influence the choice of the modification strategy; however, our findings did not show such effects. Only liquids were found to behave according to our hypothesis, and this may turn out to be due to the acceptability of dialectal variation. Furthermore, our current findings do not support the hypothesis that Spanish employs deletion as a main strategy in perception. This leads to the suggestion that perceptual grammars are not identical to production grammars. Finally, our findings suggest that linguistic proficiency may influence the choice of modification strategy as our non-native samples had slightly, but significantly, higher rate of deletion than native English speakers.

Overall, these findings provide evidence in favor of the presence of the Recoverability Principle in adult grammars since all examined groups were found to favor epenthesis over deletion. However, we could argue that these findings could be also explained by an overall preference for bisyllabic forms (Wang, 1995). That is, since the stimuli used consisted of only monosyllabic words, epenthesis would result in two syllable words. Because our current study only tested monosyllabic words, we cannot be certain that listeners have a preference for bisyllabic words. A future study could include stimuli containing bisyllabic words. Such an experiment could reveal whether listeners have a preference for epenthesis or for bisyllabicity.

Appendix. Stimuli Words Used in the Study

See Figs. 2 and 3.

Fig. 2 Experimental stimuli words

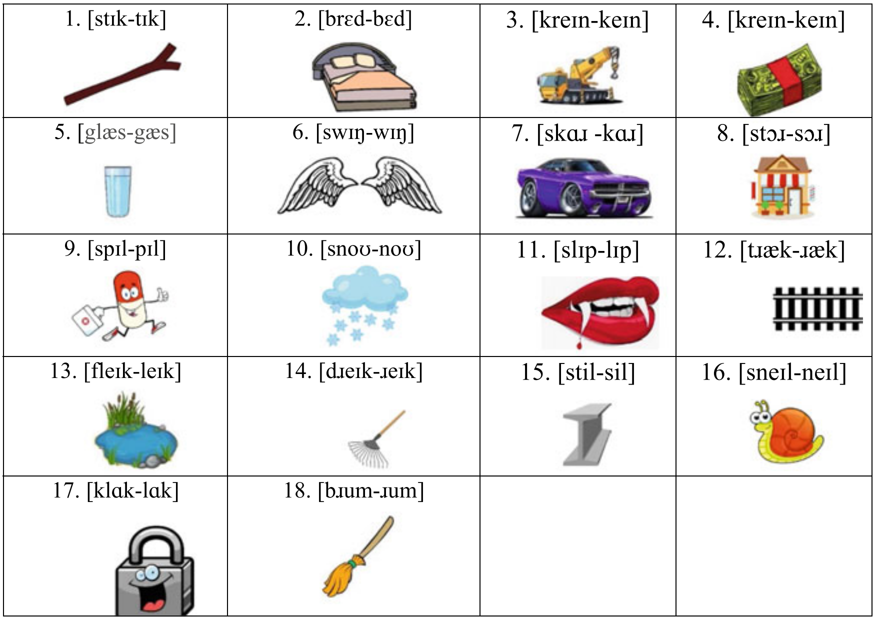

Fig. 3 Fillers

References

Abrahamsson, N. (2003). Development and recoverability of L2 codas: A longitudinal study of Chinese-Swedish interphonology. *Studies in Second Language Acquisition, 25*(3), 313–349. https://doi.org/10.1017/S0272263103000147

Benson, B. (1988). Universal preference for the open syllable as an independent process in interlanguage phonology. *Language Learning, 38*(2), 221–242. https://doi.org/10.1111/j.1467-1770.1988.tb00410.x

Bolinger, D. L. (1962). Binomials and pitch accent. *Lingua, 11*, 34–44. https://doi.org/10.1016/0024-3841(62)90011-6

Boudaoud, M., & Cardoso, W. (2009). Vocalic [e] epenthesis and variation in Farsi-English interlanguage speech. *Concordia Working Papers in Applied Linguistics*, *2*, 1–34. http://doe.concordia.ca/copal/documents/1_boudaoud_cardoso.pdf

Broselow, E., & Finer, D. (1991). Parameter setting in second language phonology and syntax. *Second Language Research, 7*(1), 35–59. https://doi.org/10.1177/026765839100700102

Cardona, G. (1988). *Pāṇini, his work and its traditions: Background and introduction* (Vol. 1). Motilal Banarsidass.

Carlisle, R. S. (2001). Syllable structure universals and second language acquisition. *International Journal of English Studies*, *1*(1), 1–19. http://revistas.um.es/ijes/article/view/47581

Clements, G. N. (1990). The role of the sonority cycle in core syllabification. In J. Kingston & M. E. Beckman (Eds.), *Papers in laboratory phonology* (pp. 283–333). Cambridge University Press. https://doi.org/10.1017/CBO9780511627736.017

Datalab.cc. (n.d.). *Jamovi* [Computer software]. https://www.jamovi.org/

de Lacy, P. V. (2002). *The formal expression of markedness* (Publication No. 3068550) [Doctoral dissertation, University of Massachusetts at Amherst]. ProQuest Dissertations Publishing.

Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., & Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance, 25*(6), 1568–1578. https://doi.org/10.1037/0096-1523.25.6.1568

Eckman, F. R. (1991). The structural conformity hypothesis and the acquisition of consonant clusters in the interlanguage of ESL learners. *Studies in Second Language Acquisition, 13*(1), 23–41. https://doi.org/10.1017/S0272263100009700

Edge, B. A. (1991). The production of word-final voiced obstruents in English by L1 speakers of Japanese and Cantonese. *Studies in Second Language Acquisition, 13*(3), 377–393. https://doi.org/10.1017/S0272263100010032

Goldsmith, J. A. (1995). *Autosegmental and metrical phonology*. Basil Blackwell.

Gregová, R. (2010). A comparative analysis of consonant clusters in English and in Slovak. *Bulletin of the Transilvania University of Brasov*, *3*(52), 79–84. Retrieved January 25, 2022, from https://www.semanticscholar.org/paper/A-COMPARATIVE-ANALYSIS-OF-CONSONANT-CLUSTERS-IN-AND-Gregov%C3%A1/674ac4c18d4633c6d5b2e36090a7240bf903cb24

Hansen, J. G. (2001). Linguistic constraints on the acquisition of English syllable codas by native speakers of Mandarin Chinese. *Applied Linguistics, 22*(3), 338–365. https://doi.org/10.1093/applin/22.3.338

Hansen, J. G. (2004). Developmental sequences in the acquisition of English L2 syllable codas. *Studies in Second Language Acquisition, 26*(1), 85–124. https://doi.org/10.1017/S0272263104026142

Hume, E. (2004). Markedness: A predictability-based approach. In M. Ettlinger, N. Fleisher, & M. Park-Doob (Eds.), *Proceedings of the Annual Meeting of the Berkeley Linguistics Society* (pp. 182–198). Berkeley Linguistics Society. https://doi.org/10.3765/bls.v30i1.948

Katamba, F. (1989). *An introduction to phonology* (Vol. 48). Longman.

Kar, S. (2010). *Syllable structure of Bangla: An optimality-theoretic approach.* Cambridge Scholars Pub.

Lombardi, L. (2003). Second language data and constraints on manner: Explaining substitutions for the English interdentals. *Second Language Research, 19*(3), 225–250. https://doi.org/10.1177/026765830301900304

Morelli, F. (2003). The relative harmony of /s+ stop/onsets obstruent clusters and the sonority sequencing. In Féry, C., & van de Vijver, R. (Eds.). *The syllable in optimality theory* (pp. 356–371). Cambridge University Press. https://doi.org/10.1017/CBO9780511497926

Núñez-Cedeño, R. A. N., Colina, S., & Bradley, T. G. (Eds.). (2014). (2nd ed.). *Fonología generativa contemporánea de la lengua española.* Georgetown University Press.

Osburne, A. G. (1996). Final cluster reduction in English L2 speech: A case study of a Vietnamese speaker. *Applied Linguistics, 17*(2), 164–181. https://doi.org/10.1093/applin/17.2.164

Parker, S. G. (2002). *Quantifying the sonority hierarchy* (Publication No. 3056268) [Doctoral dissertation, University of Massachusetts at Amherst]. ProQuest Dissertations Publishing.

Puppel, S., & Fisiak, J. (1992). The sonority hierarchy in a source-filter dependency framework. In J. Fisiak, and S. Puppel (Eds.), *Phonological investigations* (Vol. 38, pp. 467–483). John Benjamins. https://doi.org/10.1075/llsee.38.19pup

Sato, C. (1984). Phonological processes in second language acquisition: Another look at interlanguage syllable structure. *Language Learning, 34*, 43–57. https://doi.org/10.1111/j.1467-1770.1984.tb00351.x

Tropf, H. (1986). Sonority as a variability factor in second language phonology. In A. R. James & J. Leather (Eds.), *Sound patterns in second language acquisition* (pp. 173–192). De Gruyter Mouton. https://doi.org/10.1515/9783110878486-011

Tsuchida, A. (1995). English loans in Japanese: Constraints in loanword phonology. *Working Papers of the Cornell Phonetics Laboratory*, *10*, 145–164.

Wang, C. (1995). *The acquisition of English word-final obstruents by Chinese speakers* (Publication No. 9605279) [Doctoral dissertation, State University of New York at Stony Brook]. ProQuest Dissertations Publishing.

Weinberger, S. H. (1987). The influence of linguistic context on syllable simplification. In G. L. Ioup & S. Weinberger (Eds.), *Interlanguage phonology* (pp. 401–417). Newbury House Publishers.

Weinberger, S. H. (1994). Functional and phonetic constraints on second language phonology. In M. Yavaş (Ed.), *First and second language phonology* (pp. 283–302). Singular Publishing Group.

Yavaş, M. (2011). The role of sonority in the acquisition of interlanguage coda clusters. In M. Wrembel, M., Kul, M., & Dziubalska-Kolaczyk, K. (Eds.), *Achievements and perspectives in SLA of speech: New Sounds 2010* (pp. 297–307). Peter Lang. https://www.peterlang.com/document/1107711

Ali S. Alelaiwi is Assistant Professor in the Department of English at the University of Najran, Saudi Arabia. He is also a part-time English language consultant at INTO George Mason University, USA. He earned his Ph.D. in Linguistics from the Department of English at George Mason University. His research focuses on investigating the perception and production of non-native structures and examining language from an evolutionary perspective. His research interests also include foreign accents, bilingualism, and Optimality Theory.

Steven H. Weinberger is Associate Professor of Linguistics in the Department of English at George Mason University, USA. He earned his Ph.D. in Linguistics in 1988 from the University of Washington in Seattle and has taught at George Mason since 1989. He teaches graduate and undergraduate level courses in phonetics, phonology, second language acquisition, and exceptional phonologies. His research deals with language sound systems, exceptional phonologies, adult second language learning, and foreign accents. He is co-editor of *Interlanguage Phonology* (1987), and is the principal investigator and administrator of the *Speech Accent Archive.*

The Role of Plosive Codas: Recognition and Perception by Lithuanian Learners of English

Lina Bikelienė

Abstract This chapter investigates the role of voicing in English plosives in the coda position. Two tests were used to investigate this role from different perspectives: recognition and perception. Though vowel sounds are typically described according to three main factors (the front-back and high-low dimensions, and the lip aperture type), many languages, including English, distinguish a variable of duration. A direct correlation between vowel length and its phonological context has been reported in linguistic literature. The participants were 78 Lithuanian learners of English pursuing undergraduate degrees at a university in Lithuania. The recognition test aimed at addressing the phenomenon known as voicing effect in one-syllable CVC words with a plosive coda. English plosives, though perceived as voiced, are devoiced in final position. This might cause problems for Lithuanian learners of English since the correlation mark of the Lithuanian plosives is voicing. Using a variationist approach, the perception test aimed at checking the role of the English variety (British English or American English), the force of articulation, and the preceding phonological context on the perception of post-vocalic plosives. The findings indicate the Lithuanian learners' low awareness of vowel length dependence on its context, and different roles played by the studied variables.

Keywords Lithuanian learners · Voicing effect · Plosive codas · English varieties

1 Introduction

A large body of literature has investigated learner language from various perspectives and under different terms (Ellis, 1982). In the 1970s, in addition to distinguishing the target language (TL) from source languages, Nemser (1971) identified "the deviant linguistic system actually employed by the learner attempting to utilize the target language," which he named "an approximative system" (p. 115). Selinker (1972)

L. Bikelienė (✉)
Vilnius University, Vilnius, Lithuania

V. G. Sardegna and A. Jarosz (eds.), *Theoretical and Practical Developments in English Speech Assessment, Research, and Training*, Second Language Learning and Teaching,
https://doi.org/10.1007/978-3-030-98218-8_9

coined a nowadays well-established term ‘interlanguage’ (IL) and defined it as “a separate linguistic system... which results from a learner’s attempted production of a TL norm” (p. 214). In Corder’s (1981) classification, IL falls under one of the classes of ‘idiosyncratic dialects’ or, due to its unstable nature, it could also be referred to as a ‘transitional dialect.’ The view that IL is systematic but to a certain degree defective has prevailed in the linguistic literature. Bialystok and Sharwood Smith (1985) define it as “the systematic language performance... by second-language learners who have not achieved sufficient levels of analysis of linguistic knowledge or control of processing to be identified completely with native speakers” (p. 116). The scope of IL is sometimes narrowed to encompass adult learners’ linguistic system only (Tarone, 2018). Irrespective of the scope, the development of IL is believed to happen through the lects: from basilect to acrolect; the higher in language proficiency the learner is, the nearer the learner’s IL is to the TL and the better the learner’s interpersonal understanding is (Davies, 1989; Wang & Wu, 2001). The key point is that, despite individual variables, learner language is systematic (to a varying degree, though) and allows some pattern detection.

Learners’ ultimate language learning goal does not always coincide with the one formally set for university language-related courses. Their goal is believed to be “that variety of the target language which enjoys the greatest prestige and which, therefore, represents the most worthwhile investment” (Valdman, 1989, p. 276). In pronunciation classes, students often display overt preference for one of the two main English varieties: British English (BrE) and American English (AmE). Interestingly, the learners’ expressed preference is sometimes found not to be in accordance with the variety they use (Bikelienė, 2015; Smakman, 2017). Also, acquiring the TL phonology is not a straightforward process as L1 transfer is more likely to happen for learners with lower phonological abilities (see Ellis, 2015). However, the focus of this chapter is not on correlations between learners’ phonological abilities and L1 transfer but rather on examining two distinct aspects of IL: the role of phonological context and English variety (BrE or AmE). Based on prior research (Rindal & Piercy, 2013; Smakman, 2017), European learners typically use a blend of BrE and AmE, with different varieties in the dominant position. The analysis of the role BrE and AmE play in Lithuanian learners’ perception of plosive codas is expected to add one more piece to the Lithuanian English puzzle.

Given some of the similarities between plosive codas in English and Lithuanian, we can hypothesise that Lithuanian learners are likely to be able to recognise vowel duration as a voicing cue. Yet, learners with different mother tongue backgrounds are known to struggle with voicing distinction, which might require reformulating the original hypothesis. Results can be expected to provide some insights into Lithuanian English and help to target Lithuanian English-tailored pronunciation teaching. The chapter starts with background information regarding theoretical constructs for the study of IL and characteristics of the phonological features under study, and then reports and discusses the findings and pedagogical implications of this investigation.

2 Literature Review

Even though this chapter does not overtly address the question of transfer, this section briefly describes prior research on L1 transfer and markedness theory and then compares the phonological features of plosive consonants and the vowels preceding them in English and Lithuanian in order to provide the necessary background for the analysis of the effects of voicing on the preceding phonological context of post-vocalic plosives in Lithuanian accented English.

2.1 L1 Transfer and Markedness Theory

In IL studies, it is impossible to disregard the influence of the native language phonological system (see Ellis, 2015; Wang & Wu, 2001). Attempts at analysing IL concentrate around two pivotal concepts: L1 transfer and markedness theory. L1 transfer—that is, the incorporation of learners' L1 structures into the TL system—is believed to play a crucial role in the acquisition of learners' TL at all stages and all aspects of IL development (see Ellis, 2015). Findings from investigations of markedness theory, however, are not univocal. While White (1987) proposed the idea that "learners do not necessarily make a distinction between marked and unmarked structures" (p. 278), Ellis (2015) argues that marked features cause difficulties for language learners. For example, as Eckman (1977) shows, German learners of English face problems with English marked usage of voiced codas in a word-final position, whereas unmarked (or less marked) voiceless word-final consonants do not pose difficulties for English learners of German.

Correlations between a vowel's duration and its context have been reported in a number of previous studies. The findings indicate two opposing points of view. On the one hand, due to its widespread nature, vowel duration is seen as universal and unmarked (Embarki, 2016; Yoneyama & Kitahara, 2014) or 'automatic' (Ko, 2007). To support the idea of universality, Embarki (2016) reports a positive correlation of vowel duration and post-vocalic consonant voicing in male speakers' production from four Arab countries. He makes an observation of a two-fold effect: vowel lengthening before a lenis consonant and shortening before a fortis consonant. The correlation is not limited to adult speakers. Yoneyama and Kitahara (2014) conclude the existence of the universal basis of the correlation based on the analysis of Japanese infant and adult speech corpora as well as Japanese L2 learners' data. Ko's (2007) study pinpoints the age of two as the age by which children (speakers of American English) master the control of vowel length and voicing correlation. On the other hand, a number of linguists report exceptions and suggest the phenomenon of vowel length and post-vocalic consonant voicing to be of language-specific rather than universal nature (Gandour et al., 1980; Mitleb, 1984). For example, Mitleb (1984), contrary to Embarki (2016), does not observe the correlation in Arab speakers' data. Similarly, Gandour et al.'s (1980) comparison of 'normal' and laryngectomised males' production yielded no significant differences and was seen as a proof of a phenomenon being governed by English.

2.2 Plosive Consonants in English and Lithuanian

Both Lithuanian and English are equipped with plosive consonants and distinguish between voiced and voiceless sounds. Lithuanian learners of English, therefore, do not expect to have any problems in perceiving or producing English plosive stops. However, contrary to learners' expectations, the similarity of these sounds in the two languages (i.e., their L1 and TL) could have a negative effect on the correct acquisition of the TL (Ellis, 2015). It is believed that when learners do not find any equivalent of a TL sound in their L1 repertoire, they are more likely to learn its set of features than when the TL and L1 sounds share some (but not all) of their features (see Flege, 1987). In other words, the similarity of sounds in both languages can lead to creating a false sense of security and to transferring L1 features to the TL. For example, Lithuanian learners of English, even at the university level, often fail to aspirate English plosives in a word-initial position due to the wrong assumption that plosive sounds have all the same properties in both languages.

Indeed, with respect to voicing considerations, plosives in English and Lithuanian have some important similarities as well as differences. Just like in English, Lithuanian plosive consonants form a voicing correlation—/p/:/b/, /t/:/d/, and /k/:/g/—(Girdenis, 2014) and differ in one distinctive feature (voicing) (Pakerys, 1995). Lithuanian stops in a word-final position are usually unvoiced as a result of the *Law of Neutralization* of a phonemic contrast, e.g., visa/d/a—visa/t/ (Engl. always) (Pakerys, 1995). The law is, however, often violated due to its zero realisation in orthography and a scarcity of such words. In English, word-final lenis plosives, /b, d, g/, are voiceless, and word-final fortis plosives, /p, t, k/, "may have no audible release" (Cruttenden, 2014, p. 164)—that is, the fortis plosives lose their aspiration in final position. Lithuanian /p/, /t/, and /k/, on the contrary, can be aspirated solely in a word-final position (Aprijaskytė-Valdšteinienė, 1960; Pakerys, 1995). An aspirated Lithuanian plosive in final position, however, can easily alternate with an unaspirated allophone without any changes in meaning (Girdenis, 2014).

According to the place of articulation, /p/ and /b/ are bilabial in both languages. The articulation of /t/ and /d/, however, differs. While in English this pair of plosives is produced with the tip of the tongue touching the alveolar ridge, in the Lithuanian language the tongue should touch not only the alveolar ridge but the teeth as well. Therefore, the consonants /t/ and /d/ are classed as alveolar in English, and dental (Kazlauskienė, 2018; Urbanavičienė, 2019; Urbanavičienė et al., 2019), apical dental (Girdenis, 2014) or dental/alveolar (Kushnir, 2016) in Lithuanian. /k, g/ in English are labelled as velar. In Lithuanian, /k, g/ are categorised as non-apical (Girdenis, 2014), velar (Kazlauskienė, 2018; Urbanavičienė et al., 2019) or guttural (Bacevičiūtė, 2009) plosives. A slight difference in articulation, though not sufficient to cause recognition problems, should be noticed. In the production of the Lithuanian velar plosives, the tongue should be pressed to the back of the hard palate—that is, they are pronounced between the back of the tongue and the soft palate and not with the tongue pressed to the soft palate like in English.

2.3 Vowel Duration in Preceding Post-Vocalic Consonants

Both English and Lithuanian speakers differentiate vowels according to the length dimension and perceive them as longer in stressed syllables (Lunden, 2017; Pakerys, 1995). In English, the length difference, however, is believed to be not stress-related but rather depend on "vowel-intrinsic durational characteristics" (Ciszewski, 2012, p. 223). The analysis of one-syllable words highlights one more factor affecting vowel length: the voicing of a post-vocalic consonant. The existence of such a correlation has been attested in native English (see Cho, 2016, for AmE, BrE, and New Zealand English; Holt et al., 2016, for African American; Tanner et al., 2019, for BrE and North AmE; Tauberer & Evanini, 2009, for BrE dialects) and English IL (see Bikelienė & Vaitkevičiūtė, 2018, for Lithuanian English; Chung, 2019, Park et al., 2019, for Korean English; Reinisch & Penney, 2019, for German English; Skarnitzl & Šturm, 2016, for Czech English), as well as in a number of other languages such as in Brazilian Portuguese (Alves & Brisolara, 2020), Georgian (Beguš, 2017), German (Zihlmann, 2020), Hindi (Sanker, 2019), Lithuanian (Campos-Astorkiza, 2012), Nepali (Schwarz, 2018), Arabic (Fathi & Qassim, 2020), Italian and Polish (Coretta, 2019).

There is no unanimous agreement on the treatment of the voicing effect in the linguistic literature. Is it articulatory motivated or speaker controllable (cf. Ciszewski, 2012)? In what way are the preceding vowels affected by a post-vocalic consonant? This uncertainty is well attested in the existent terminology. 'Pre-fortis clipping' (Wells, 1990) or 'shortening' (Cruttenden, 2014) suggest a shortening rule. 'Vowel-length effect' (Ko, 2007), 'post-vocalic consonant voicing effect' (Tauberer & Evanini, 2009), 'consonantal voicing effect' (Beller-Marino, 2014), and 'voicing effect' (Yoneyama & Kitahara, 2014), on the other hand, are not overtly specific and may include studies reporting a lengthening rule (Gandour et al., 1980; Scheer, 2017; Tauberer & Evanini, 2009). Since this chapter is of a descriptive nature, one of the non-directional terms, the 'voicing effect' will be used.

3 The Study

The main goal of this study was to gain some insights into the understanding of one aspect of Lithuanian English by checking whether Lithuanian learners of English are aware of the correlation between the voicing of plosive codas and the length of the preceding vowel. Specifically, the following research questions guided this research:

- Do Lithuanian learners of English acknowledge the existence of a different vowel length concept based on the voicing of plosive codas?
- Do the preceding vowel quality, the place of coda articulation, and the English language variety (BrE and AmE) play a role in the perception of plosive codas?

4 Methods

4.1 Participants

The participants were 78 first-year students ($F = 72$; $M = 6$; aged 18–19), majoring in English and another language (French/Norwegian/Russian/Spanish), from the Faculty of Philology at Vilnius University in Lithuania. All the students were native speakers of Lithuanian with a B1-B2 level of English (according to the State level Matura English Examination) (Nacionalinis egzaminų centras, n.d.). The participants were starting their introductory course of English phonetics and had no theoretical knowledge related to the voicing effect.

4.2 Data Collection and Analysis

Two tests were designed for this study: a recognition test and a perception test. For ecological validity, the data was obtained in an everyday classroom environment using standard classroom procedures. The students were informed that the tests would not be graded and the results would be used for research purposes to eliminate any possible influence of a stress factor.

During the English phonetics course, students mainly focused on the standard or a neutral type of accent for British English with minor attention paid to General American English. There is no consistency in terminology regarding the former. It can appear under the terms of non-regional pronunciation (NRP) (Collins & Mees, 2013), General British (GB) (Cruttenden, 2014), Standard Southern British (SSB) (Lindsey, 2019) or Standard Southern British English Pronunciation (SSBE) (Knight, 2012). For the tests, *Cambridge English Pronouncing Dictionary 18th edition* and *Cambridge Dictionary Online* were used; thus, the chapter follows their tradition and refers to the two main standard varieties as British English (BrE) and American English (AmE).

Test 1: Recognition of vowels. The students were provided with 28 minimal pairs of CVC words with a plosive consonant in the coda position (see minimal pairs in Appendix). All the minimal pairs were provided in an orthographic form arranged randomly. The minimal pair *hark—Hag* was supplemented with a transcription for the element *Hag* due to its irregular pronunciation. The task asked for a decision whether the vowels of each minimal pair in the BrE variety were of the same length or different. In the case of the latter, the students were requested to indicate the word with a longer vowel.

Test 2: Perception of plosive codas. The students' perception of the voicing plosive in the coda position was tested with an auditory test. The recordings of one-syllable CVC minimal pair words (where possible) in BrE and AmE were played in random order. It is important to note that the minimal pairs were formed based on BrE. For each phonological context (/? ɪ _ /, /? e _ /, /? ʌ _ /, /? ɒ _ /, /? ɑ: _ /, /? ɔ: _ /, /? ɜ: _ /, /? æ _ /, /? i: _ /, /and /? u: _ /), there were six words in two varieties per line, that is, 12 sounds to be inserted in the table. Since there were no minimal pairs for the /? ʊ _ / context, the students listened to the recordings of three individual words in the two varieties under consideration.

For analysis purposes, BrE target vowels are referred to by the lexical set keywords proposed by Wells (1982): /ɪ/ = KIT, /e/ = DRESS, /æ/ = TRAP, /ʌ/ = STRUT, /ʊ/ = FOOT, /ɑ:/ = BATH (together with START), /ɒ/ = CLOTH, /ɜ:/ = NURSE, /i:/ = FLEECE, /ɔ:/ = THOUGHT (together with FORCE), and /u:/ = GOOSE. The research methodology was based on a variationist approach Level II (two linguistic factors) and Level I (linguistic and social factors) (Hazen, 2017). The relationship between the force of articulation of the coda plosive, the features of the preconsonantal vowel, the variety of English, and the voicing effect are presented with the help of descriptive statistics. Test 2 results were evaluated with the Wilcoxon non-parametric test at the 0.05 significance level.

5 Results

5.1 Test 1: Recognition of Vowels

The results of the recognition test indicate participants' general tendency (73%) to perceive vowels in any given minimal pair as having the same length regardless of the following consonant. Even though Lithuanian vowels can be short and long depending on the following consonant (Campos-Astorkiza, 2012) and the same letter can be rendered as a short (e.g., *mes* Engl. *will throw*) and long vowel (e.g., *mes* Engl. *we*), the large majority of the learners did not seem to believe that a vowel phoneme in English could be of two different lengths. The results, however, are promising since, when the learners recognised a durational difference, they indicated a longer sound preceding a lenis than a fortis plosive nearly five times more often (552 and 111 cases, respectively).

The comparison of plosive pairs (Table 1) indicates that in the case of the bilabial plosive codas (/p/ and /b/), no vowel length difference was observed in 80 percent of the cases ($M = 62.64$, $SD = 14.31$), with a range from 50 to 100 percent. It was followed by velar (/k/ and /g/) and alveolar (/t/ and /d/) plosive codas, with mean percentage 70 ($M = 54.67$, $SD = 13.85$) and 67 ($M = 52.18$, $SD = 11.44$), respectively.

Table 1 Vowel length recognition in word pairs with plosive codas (percentage)

	Same Length			Longer before Fortis			Longer before Lenis		
	Bilabial	Alveolar	Velar	Bilabial	Alveolar	Velar	Bilabial	Alveolar	Velar
/? ɪ _ /	92	90	92	0	1	0	8	9	8
/? e _ /	89	62	81	9	5	3	3	33	17
/? ʌ _ /	87	82	94	1	3	1	12	15	5
/? ɒ _ /	67	47	58	3	5	3	31	47	40
/? ɑ: _ /	64	60	46	4	1	9	32	39	45
/? ɔ: _ /	100	59	58	0	17	9	0	24	33
/? ɜ: _ /	100	73	77	0	1	0	0	26	23
/? æ _ /	56	46	50	4	5	15	40	49	35
/? i: _ /	78	77	76	4	0	6	18	23	18
/? ʊ _ /	100	82	NA	0	0	NA	0	18	NA
/? u: _ /	50	58	NA	33	0	NA	17	42	NA
Total	80	67	70	5	4	5	15	30	25

Both alveolar and velar plosive codas proved to be more context-related. For the former, two instances were observed, where the percentage for the same length was below 50. The LOT vowel was reported as being of the same length with the same frequency as being longer before a lenis member of the pair (47 percent of cases). The TRAP vowel was marked as being of the same length in both minimal pair words in approximately the same number as in LOT words (46 percent of cases). In the minimal pair words ending in /k/ or /g/, only BATH vowel was marked as being of the same length in less than half of the cases (46 percent). Such results indicate the need for students to be exposed to finer features of the BrE sounds.

5.1.1 The Influence of Plosive Codas on Vowels According to the Front/Back Dimension

As Fig. 1 shows, the highest mean percentage for correct length recognition (28%) ($M = 26.09$, $SD = 8.12$) can be observed when a plosive coda follows one of the back vowels (e.g., as in /? u: _ /, /? ʊ _ /, /? ɔ: _ /, /? ɒ _ /, and /? ɑ: _ /). In such a phonological context, the learners failed to notice any difference in length in 65 percent of their choices. The lowest mean percentage for correct length recognition was when a plosive is preceded by a central vowel /ɜː/ or /ʌ/ (14 percent of answers choosing lenis, $M = 10.5$, $SD = 7.79$, while 86 percent indicating no difference). The results for front vowels (FLEECE, KIT, DRESS, and TRAP) show 22 percent of correct answers ($M = 16.83$, $SD = 11.36$).

The analysis of plosive pairs according to the place of articulation highlights a general tendency for bilabial plosive codas (/b/ and /p/) to trigger the lowest number

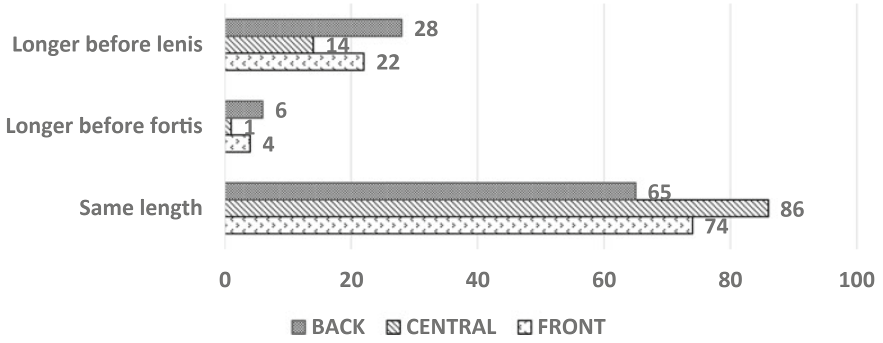

Fig. 1 Vowel length recognition according to front/back dimension in word pairs with plosive codas (percentage)

of correct answers preceded by all the vowel groups, front (17 percent), central (6 percent), and back (16 percent). Interestingly, the back vowels are relatively the easiest to recognise when followed by both velar and alveolar plosives (39 and 34 percent, respectively). For the central vowels, the durational accuracy rate was three and a half times higher when they were followed by the alveolar and more than twice higher by the velar than by the bilabial consonant.

5.1.2 The Influence of Plosive Codas on Vowels According to the Close/Open Dimension

The highest mean percentage accuracy (33) was observed when a plosive coda followed a (mid)open vowel (/æ/, /ʌ/, /ɒ/, /ɑː/) ($M = 25.25$, $SD = 11.2$), followed by a (mid)close vowel (/iː/, / ɪ/, /uː/, /ʊ/) (16 percent) ($M = 13.89$, $SD = 8.33$) and a mid vowel (/e/, /ɜː/, /ɔː/) (14 percent) ($M = 13.78$, $SD = 10.62$) (Table 2).

As Table 2 also shows, the easiest phonological context for the learners appeared to be a (mid)open vowel followed by an alveolar plosive (38 percent of correct answers) as in *hat*. Nearly every third correct answer was when a (mid)open vowel was followed by a velar (as in *duck*) or a bilabial plosive (as in *cop*) (31 and 29 percent, respectively). The most controversial results can be witnessed in the

Table 2 Vowel length recognition according to close/open dimension in word pairs with plosive codas (percentage)

	Same length			Longer before fortis			Longer before lenis		
	Bilabial	Alveolar	Velar	Bilabial	Alveolar	Velar	Bilabial	Alveolar	Velar
(MID)CLOSE	80	77	84	9	0.3	3	11	23	13
MID	96	65	72	3	7	4	1	28	24
(MID)OPEN	69	59	62	3	4	7	29	38	31

Table 3 Vowel length recognition according to short/long dimension in word pairs with plosive codas (percentage)

	Same length			Longer before fortis			Longer before lenis		
	Bilabial	Alveolar	Velar	Bilabial	Alveolar	Velar	Bilabial	Alveolar	Velar
SHORT	82	67	75	3	3	4	15	29	21
LONG	79	65	64	8	4	6	13	31	30

phonological context with a mid vowel. While the accuracy rate is relatively high for alveolar (28 percent) and velar plosives (24 percent), only one percent of correct answers was observed when such a vowel was followed by the bilabial plosive.

5.1.3 The Influence of Plosive Codas on the Preceding Vowel According to the Short/Long Dimension

The rate of correct answers is only insignificantly higher when a plosive is preceded by a short vowel ($M = 17.94$, $SD = 12.38$) than by a long vowel ($M = 22.08$, $SD = 7.59$) (22 and 25 percent, respectively). As with the other dimensions, the lowest accuracy rate was with bilabial plosive codas: 13 percent after a long and 15 percent after a short vowel (Table 3).

The recognition of correct vowel length for short and long vowels before an alveolar plosive differs only by two percent (29 and 31 percent, respectively). The Lithuanian learners showed the biggest differences in recognizing the longer sound when it was followed by a velar plosive: 21 percent following a short and 30 percent following a long vowel.

5.2 Test 2: Perception of Plosive Codas

The results of Test 2 show the Lithuanian learners' ability to better discriminate the coda voicing in AmE than in BrE. In 89 and 84 percent of instances, respectively, the correct voicing was indicated even if the coda itself was marked incorrectly. The difference is statistically significant at $p < 0.05$ ($z = 2.1028$, $p = 0.03572$). The analysis of the students' perception of fortis and lenis codas separately, however, yielded different results. No statistically significant differences were observed in either the comparison of fortis and lenis codas separately in both BrE and AmE ($W = 16$, the critical value is 8, $p > 0.05$ for fortis codas; $W = 12.5$, the critical value is 10, $p > 0.05$ for lenis codas), or between fortis and lenis codas in each variety separately ($W = 26.5$, the critical value is 10, $p > 0.05$ for BrE; $W = 30$, the critical value is 10, $p > 0.05$ for AmE).

The comparison of the perception of correct coda voicing in different phonological contexts indicates two cases when learners found BrE more complicated than AmE.

When the coda was preceded by a FOOT vowel, as in /p ʊ _/, in BrE, the students tended to hear a fortis consonant /t/ instead of a lenis consonant /d/. This mistake was made in 82 percent of all the cases. Interestingly, in AmE, this happened in only 14 percent of all the cases. The other problematic phonological context in BrE was when a lenis consonant (/b/, /d/ or /g/) followed a KIT vowel, as in *rib*, *kid*, and *pig*. The error rate was 44 percent in BrE while in AmE it was only 8 percent. Neither of the analysed contexts posed problems when the coda was fortis as in *rip*, *hat*, *cop*. The error rate varied from 2 to 34 percent for BrE and from 3 to 20 percent for AmE.

Interestingly, the results for students' perception of correct coda indicate the same tendency irrespective of the English variety under consideration. The Lithuanian students more often tended to perceive correctly fortis than lenis plosive codas: 82 and 70 percent in BrE, and 83 and 75 in AmE. The differences, however, are not statistically significant: $W = 13$, the critical value is 10, $p > 0.05$ for BrE; $W = 17$, the critical value is 10, $p > 0.05$ for AmE.

According to the front-back dimension of the preceding vowel, no significant differences were observed in the students' perception of correct plosive codas. In both varieties, when the central vowel was followed by a plosive, as in /? ɜː _ / or /? ʌ _ /, the accuracy rate was nearly identical regardless of the coda voicing (78 and 82 percent for BrE, and 85 and 84 in AmE, respectively). In the case of back vowels, in both varieties, the difference in the accuracy rate was nearly identical (83 vs. 66 and 84 vs. 66 percent). In BrE, when the plosive coda was preceded by a front vowel, students perceived the correct coda with almost the same accuracy as in the case of a back vowel. In AmE, however, fortis and lenis consonants were perceived with the accuracy of 80 and 81 percent.

In sum, the phonological context, according to the high/low and short/long dimensions of the preceding vowel, proved to have no influence on the perception of the correct coda. The only exception was observed in the case of (mid)high vowels in BrE, where the accuracy rate for the perception of lenis codas was only 48 percent.

6 Discussion

The findings of the recognition test provide strong evidence indicating that Lithuanian learners of English face difficulties using vowel duration as a cue to the voicing of the following consonant. These findings contradict the results reported in Bikelienė and Vaitkevičūtė (2018) regarding the production of checked unrounded vowels in CVC words with fortis and lenis codas. According to Campos-Astorkiza (2012), Lithuanian speakers should be familiar with vowel duration as a cue to voicing a following consonant from their native language. Arguably, native language knowledge is used automatically and, thus, should be more likely to manifest itself in tasks that do not require conscious efforts to implement. The recognition test, however, revealed that the learners had problems transferring their L1 features to English. The findings of the recognition test are in line with other studies on voicing effect in learner English with different mother tongue backgrounds (e.g., Reinisch & Penney,

2019) that show the difficulty to use the vowel length cue. Though both English and Lithuanian devoice final consonants, the ratio between fortis and lenis plosive codas is in sharp contrast in the two languages. Due to the Lithuanian language morphemic structure, word final plosives are mainly restricted to fortis consonants potentially making Lithuanian learners to treat word final lenis plosives as *new* and *difficult* (Reinisch & Penney, 2019) and, according to the Markedness Differential Hypothesis (Eckman, 1977), likely also causing problems in distinguishing coda voicing in English.

The perception test suggested a slight tendency (even though statistically insignificant) for Lithuanian learners of English to accurately perceive fortis English plosives more often than their lenis counterparts. This finding is in line with Reinisch and Penney's (2019) argument indicating that fortis consonants were *easier* for learners. Interestingly, the same tendency was partially observed for the recognition test: When learners recognised vowel durational difference, nearly five times more often they correctly indicated a vowel to be longer before a lenis than a fortis plosive. The results suggest that Lithuanian learners are able, to a certain extent, to transfer L1 knowledge regarding the phenomenon of voicing effect (Campos-Astorkiza, 2012) to TL. The statistically insignificant difference between correctly and incorrectly perceived plosive codas for the perception test and overall low numbers of correct answers for the recognition test, on the other hand, point towards the need of explicit instruction, as understanding phonetic features is important for phonetic accuracy and may have an impact on intelligibility (Levis, 2018). To make a stronger claim on Lithuanian learners' understanding of vowel duration as a cue to coda voicing, the tests could be repeated on other student groups and at different times of their studies, that is pre- and post-introduction to the phenomenon in English.

Finally, the analysis of a phonological context indicates the results to be not homogeneous: bilabial plosive codas are more likely to be problematic for Lithuanian learners of English than alveolar or velar plosives. Also, the examination of the role of the English variety variable for vowel length-coda voicing correlation seems to support a blended-variety idea (cf. Rindal & Piercy, 2013; Smakman, 2017). The evidence suggests that while AmE seems an 'easier' variety to hear voicing, no significant differences between BrE and AmE were observed in perceiving the correct coda sounds.

7 Implications

This chapter attempted to describe some voicing effect-related features in Lithuanian learner English phonology. Vocalic duration is known to be not the only and sometimes even not the most crucial factor affecting voicing decisions (Nittrouer, 2004); thus, a more detailed study would benefit a better understanding of the phenomenon. It would be reasonable alongside recognition and perception to perform a thorough articulatory analysis of Lithuanian learners' production of fortis/lenis plosive codas in different phonological environments. It could be hypothesised that raised awareness of vowel duration as a cue for voicing could be an important factor in enhancing listening accuracy.

To eliminate the familiarity effect, the tests could be replicated with non-existent words. This could contribute to the establishment of the place of Lithuanian English in the IL continuum across different learner populations. A better understanding of the differences between Lithuanian English and native English phonology could help the enhancement of teaching and the design of teaching materials oriented at learners with a particular mother tongue.

8 Conclusions

It has been previously reported that learners' ability to produce sounds follow their ability to recognise them (Ellis, 2015). Based on this claim, the chapter set out to investigate any potential differences in the recognition and perception of the voicing effect in Lithuanian English concentrating on the role played by the preceding vowel quality, the place of articulation of the final consonant, and the English language variety (BrE and AmE) in the perception of plosive codas. Though the voicing effect is sometimes referred to as universal, the results of the analysis are in accordance with a large body of IL studies, which indicate low awareness of the link between the coda voicing and vowel length irrespective of the force of articulation of the final consonant.

The results of the study signal that perception of the coda in the listening task precedes the recognition of the phenomenon of voicing effect on the preconsonantal vowel. The findings are not surprising since the perception test required the students' listening skills, which were relatively good, while the recognition test implied at least passive knowledge of the linguistic phenomenon of voicing effect, which requires explicit teaching in the general perception of voicing only. The comparison of the perception of fortis and lenis plosive codas was indicative of an insignificantly better perception of fortis consonants. The preceding phonological context had no important effect on the learners' perception.

Appendix

Minimal pairs

rip – rib, cob – cop, beep – Beeb, dead – debt, cart – card, had – hat, moot – mood, duck – dug, baulk – Borg, league – leek, Depp – deb, carb – carp, loop – lube, mud – mutt, caught – cord, heed – heat, pig – pick, block – blog, berg – berk, pub – pup, cab – cap, kit – kid, nod – not, Birt – bird, pud – put, peg – peck, hark – Hag /hɑːg/, hag – hack.

References

Alves, U. K., & Brisolara, L. B. (2020). Listening to accented speech in Brazilian Portuguese: On the role of fricative voicing and vowel duration in the identification of /s/ – /z/ minimal pairs produced by speakers of L1 Spanish. *Journal of Portuguese Linguistics, 19*(1), 1–6. https://doi.org/10.5334/jpl.237

Aprijaskytė-Valdšteinienė, R. (1960). Anglų ir lietuvių priebalsių lyginimas. *Kalbotyra*, *2*, 167–185. https://doi.org/10.15388/Knygotyra.1960.18510

Bacevičiūtė, R. (2009). Lietuvių kalbos garsynas Augusto Schleicherio "Litauische Grammatik". *Baltistica*, *44*(2), 335–346. https://doi.org/10.15388/baltistica.44.2.1321

Beguš, G. (2017). Effects of ejective stops on preceding vowel duration. *The Journal of the Acoustical Society of America, 142*(4), 2168–2184. https://doi.org/10.1121/1.5007728

Beller-Marino, Y. (2014). *Consonantal voicing effects on vowel duration in Italian-English bilinguals* (Publication No. 3641876) [Doctoral dissertation, The City University of New York]. ProQuest Dissertations Publishing.

Bialystok, E., & Sharwood Smith, M. (1985). Interlanguage is not a state of mind: An evaluation of the construct for second-language acquisition. *Applied Linguistics, 6*(2), 101–117. https://doi.org/10.1093/applin/6.2.101

Bikelienė, L. (2015). Lithuanian Learners' English: British or American? *Verbum*, *6*, 29–40. https://doi.org/10.15388/Verb.2015.6.8806

Bikelienė, L., & Vaitkevičiūtė, M. (2018). The coda voicing contrast in Lithuanian learners' English. *Verbum*, *9*, 11–20. https://doi.org/10.15388/Verb.2018.2

Campos-Astorkiza, R. (2012). Length contrast and contextual modifications of duration in the Lithuanian vowel system. *Baltic Linguistics*, *3*, 9–41. https://doi.org/10.32798/bl.418

Cho, H. (2016). Variation in vowel duration depending on voicing in American, British, and New Zealand English. *Phonetics and Speech Sciences*, *8*(3), 11–20. https://doi.org/10.13064/KSSS.2016.8.3.011

Chung, J. (2019). Production and perception of English vowel length depending on the following consonant voicing by Korean learners of English. *The Journal of the Acoustical Society of America, 146*, 2958. https://doi.org/10.1121/1.5137276

Ciszewski, T. (2012). Stressed vowel duration and phonemic length contrast. *Research in Language, 10*(2), 215–223. https://doi.org/10.2478/v10015-011-0049-2

Collins, B., &. Mees, I. M. (2013). *Practical phonetics and phonology. A resource book for students* (3rd ed.). Routledge. https://doi.org/10.4324/9780203080023

Corder, S. P. (1981). *Error analysis and interlanguage.* Oxford University Press.

Coretta, S. (2019). An exploratory study of voicing-related differences in vowel duration as compensatory temporal adjustment in Italian and Polish. *Glossa: A Journal of General Linguistics*, *4*(1), 125. https://doi.org/10.5334/gjgl.869

Cruttenden, A. (2014). *Gimson's pronunciation of English* (8th ed.). Routledge. https://doi.org/10.4324/9780203784969

Davies, A. (1989). Is international English an interlanguage? *TESOL Quarterly, 23*(3), 447–467. https://doi.org/10.2307/3586920

Eckman, F. R. (1977). Markedness and the contrastive analysis hypothesis. *Language Learning, 27*(2), 315–330. https://doi.org/10.1111/j.1467-1770.1977.tb00124.x

Ellis, R. (1982). The origins of interlanguage. *Applied Linguistics, 3*(3), 207–223. https://doi.org/10.1093/applin/3.3.207

Ellis, R. (2015). *Understanding second language acquisition* (2nd ed.). Oxford University Press.

Embarki, M. (2016). *Voicing effects an absolute universal or language specific: New evidence from modern Arabic*. https://doi.org/10.13140/RG.2.1.1520.9208

Fathi, H. M., & Qassim, Z. R. (2020). An acoustic study of the production of Iraqi Arabic vowels. *Journal of Al-Frahedis Arts*, *12*(40), 692–704. http://dx.doi.org/https://doi.org/10.51990/2228-012-040-008

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. https://doi.org/10.1016/S0095-4470(19)30537-6
Gandour, J., Weinberg, B., & Rutkowski, D. (1980). Influence of post-vocalic consonants on vowel duration in esophageal speech. *Language and Speech, 23*(2), 149–158. https://doi.org/10.1177/002383098002300202
Girdenis, A. (2014). *Theoretical foundations of Lithuanian phonology* (2nd ed.). (S. Young, Trans.). Vilnius University. Lithuanian edition: Girdenis, A. (2003). Teoriniai lietuvių fonologijos pagrindai. Antrasis pataisytas ir papildytas leidimas. Mokslo ir enciklopedijų leidybos institutas.
Hazen, K. (2017). Language variation: Sociolinguistic variationist analysis. In M. Aronoff, & J. Rees-Miller (Eds.), *The handbook of linguistics* (2nd ed., pp. 519–540). Wiley Blackwell. https://doi.org/10.1002/9781119072256.ch25
Holt, Y. F., Jacewicz, E., & Fox, R. (2016). Temporal variation in African American English: The distinctive use of vowel duration. *Journal of Phonetics & Audiology, 2*(2), 1–8. https://doi.org/10.4172/2471-9455.1000121
Kazlauskienė, A. (2018). *Bendrinės lietuvių kalbos fonetikos ir fonologijos pagrindai.* VDU leidykla.
Knight, R. A. (2012). *Phonetics.* Cambridge University Press.
Ko, E.-S (2007). Acquisition of vowel duration in children speaking American English. In *Proceedings of Interspeech 2007* (pp. 1881–1884). International Speech Communication Association (ISCA) http://dx.doi.org/https://doi.org/10.21437/Interspeech.2007-523
Kushnir, Y. (2016). *Structure of Lithuanian.* Retrieved August 3, 2020, from https://home.uni-leipzig.de/~yuriykushnir/strucclith/class_1.pdf
Levis, J. M. (2018). *Intelligibility, oral communication, and the teaching of pronunciation.* Cambridge University Press.
Lindsey, G. (2019). *English after RP. Standard British pronunciation today*. Palgrave Macmillan. https://doi.org/10.1007/978-3-030-04357-5
Lunden, A. (2017). Duration, vowel quality, and the rhythmic pattern of English. *Laboratory Phonology, 8*(1), 1–20. https://doi.org/10.5334/labphon.37
Mitleb, F. (1984). Voicing effect on vowel duration is not an absolute universal. *Journal of Phonetics, 12*(1), 23–27. https://doi.org/10.1016/S0095-4470(19)30847-2
Nacionalinis egzaminų centras (n.d.). *Užsienio kalbos brandos egzamino mokinių pasiekimų lygių aprašas.* Retrieved July 22, 2021, from https://www.nec.lt/failai/5699_UK_APRASAS.pdf
Nemser, W. (1971). Approximative systems of foreign language learners. *IRAL: International Review of Applied Linguistics in Language Teaching, 9*(2), 115–123. https://doi.org/10.1515/iral.1971.9.2.115
Nittrouer, S. (2004). The role of temporal and dynamic signal components in the perception of syllable-final stop voicing by children and adults. *The Journal of the Acoustical Society of America, 115*(4), 1777–1790. https://doi.org/10.1121/1.1651192
Pakerys, A. (1995). *Lietuvių bendrinės kalbos fonetika: Vadovėlis aukštųjų mokyklų studentams.* Žara.
Park, W., Hyunjun, L., & Seok-Chae, L. (2019). The effect of word-final stops' voicing on the vowel duration and its relation with Korean speakers' English proficiency. *The Journal of the Acoustical Society of America, 146*(4), 28–43. https://doi.org/10.1121/1.5136860
Reinisch, E., & Penney, J. (2019). The role of vowel length and glottalization in German learners' perception of the English coda stop voicing contrast. *Laboratory Phonology: Journal of the Association for Laboratory Phonology, 10*(1), 1–26. https://doi.org/10.5334/labphon.176
Rindal, U., & Piercy, C. (2013). Being 'neutral'? English pronunciation among Norwegian learners. *World Englishes, 32*(2), 211–229. https://doi.org/10.1111/weng.12020
Sanker, Ch. (2019). Influence of coda stop features on perceived vowel duration. *Journal of Phonetics*, *75*, 43–56. https://doi.org/10.1016/j.wocn.2019.04.003
Scheer, T. (2017). Voice-induced vowel lengthening. *Papers in Historical Phonology, 2*, 116–151. https://doi.org/10.2218/pihph.2.2017.1910

Schwarz, M. (2018). Pre-coda vowel duration in Nepali. *The Journal of the Acoustical Society of America, 143*(3), 1756–1756. https://doi.org/10.1121/1.5035749
Selinker, L. (1972). Interlanguage. *IRAL: International Review of Applied Linguistics in Language Teaching*, *10*(3), 209–231. https://doi.org/10.1515/iral.1972.10.1-4.209
Skarnitzl, R., & Šturm, P. (2016). Pre-fortis shortening in Czech English: A production and reaction-time study. *Research in Language, 14*(1), 1–14. https://doi.org/10.1515/rela-2016-0005
Smakman, D. (2017). The use of native-speaker pronunciation models. In S. Lindenburg S. & D. Smakman (Eds.), *Proceedings of the Van Schools tot Scriptie III Colloquium* (pp. 97–101). Leiden University Repository. https://scholarlypublications.universiteitleiden.nl/access/item%3A3144638/view
Tanner, J., Sonderegger, M., Stuart-Smith, J., & SPADE Data Consortium. (2019). Vowel duration and the voicing effect across English dialects. *Toronto Working Papers in Linguistics*, *41*(1), 1–13. https://doi.org/10.33137/twpl.v41i1.32769
Tarone, E. (2018). Interlanguage. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1–7). John Wiley and Sons, Ltd. https://doi.org/10.1002/9781405198431.wbeal0561.pub2
Tauberer, J., & Evanini, K. (2009). Intrinsic vowel duration and the post-vocalic voicing effect: Some evidence from dialects of North American English. In *Proceedings of Interspeech* 2009 (pp. 2211–2214). International Speech Communication Association (ISCA). https://doi.org/10.21437/Interspeech.2009-629
Urbanavičienė, J. (2019). Consonants of the Lithuanian language in IPA context: Identification problems. *Сучасні тенденції фонетичних досліджень: Збірник матеріалів III Круглого столу з міжнародною участю* (19 квітня 2019 р.), 45–49. https://ela.kpi.ua/handle/123456789/35322
Urbanavičienė, J., Indričāne, I., Jaroslavienė, J., & Grigorjevs, J. (2019). *Baltų kalbų garsynas XXI a. pradžioje: Priebalsių instrumentinis tyrimas: kolektyvinė monografija*, kn. 2. Lietuvių kalbos institutas. http://lki.lt/wp-content/uploads/2020/03/Balt%C5%B3_kalb%C5%B3_garsynas_XXI_a_prad%C5%BEioje_2_knyga_2019.pdf
Valdman, A. (1989). Classroom foreign language learning and language variation. In M. R. Eisenstein (Ed.), *The dynamic interlanguage: Topics in language and linguistics* (pp. 261–277). Springer. https://doi.org/10.1007/978-1-4899-0900-8_16
Wang, H. S., & Wu, J. C. (2001). The effect of vowel duration on the perception of post-vocalic voiced/voiceless consonants. *Concentric Studies in English Literature and Linguistics*, *27*, 35–52. http://www.concentric-linguistics.url.tw/upload/articlesfs91402102220157532.pdf
Wells, J. C. (1982). *Accents of English 1: An introduction.* Cambridge University Press.
Wells, J. C. (1990). Syllabification and allophony. In S. Ramsaran (Ed.), *Studies in the pronunciation of English: A commemorative volume in honour of A.C. Gimson* (pp. 76–86). Routledge.
White, L. (1987). Markedness and second language acquisition. The question of transfer. *Studies in Second Language Acquisition*, *9*(3), 261–285. https://doi.org/10.1017/S0272263100006689
Yoneyama, K., & Kitahara, M. (2014). Voicing effect on vowel duration: corpus analyses of Japanese infants and adults, and production data of English learners. *Journal of the Phonetic Society of Japan*, *18*(1), 30–39. https://doi.org/10.24467/onseikenkyu.18.1_30
Zihlmann, U. (2020). Vowel and consonant length in four Alemannic dialects and their influence on the respective varieties of Swiss Standard German. *Wiener Linguistische Gazette, 86*, 1–46. https://doi.org/10.5167/uzh-186717

Lina Bikelienė is Associate Professor at the Institute of Foreign Languages, the Faculty of Philology at Vilnius University in Lithuania, where she teaches English phonetics, media linguistics, and corpus linguistics. Her research interests range from English phonetics and phonology and current phonological processes to learner language. Dr. Lina Bikelienė is on the editorial board of *Verbum*, an open-access journal.

Exploring How *YouGlish* Supports Learning English Word Stress: A Perception Study

Veronica G. Sardegna and Anna Jarosz

Abstract Wrong lexical stress placement affects listeners' ability to recognize words, especially when the misplacement leads to vowel quality change. Yet, despite its importance for intelligibility, EFL teachers rarely devote class time to teaching word stress. Based on suggestions for an intelligibility-based approach, this study investigated the effectiveness of using *YouGlish* (www.youglish.com) to support students' out-of-class perception practice of lexical stress. Participants were 12 Polish EFL learners (16–18 years old) taking a language course at a school in Poland. To facilitate students' learning of long academic words, the teacher assigned worksheets for homework. The worksheets explained two main orthographic rules that guide the stress of English polysyllabic words, and offered some practice applying the rules. Students were instructed to practice lexical stress using the worksheets and listening to the words pronounced in *YouGlish* out of class for four weeks. Data were gathered from pre- and post-tests assessing students' ability to predict and perceive the stress of 20 English polysyllabic words, a background questionnaire, and pronunciation trackers eliciting students' opinions on the practice materials and experience. Students' self-reported practices explained differences in students' improvement with regards to predicting and perceiving word stress in polysyllabic words. Pedagogical implications of the findings are discussed.

Keywords English word stress · Orthographic rules · Polysyllabic words · Instructional technology · Autonomous learning · EFL learners

V. G. Sardegna (✉)
Duquesne University, Pittsburgh, PA, USA
e-mail: sardegnav@duq.edu

A. Jarosz
University of Lodz, Łódź, Poland
e-mail: anna.jarosz@uni.lodz.pl

V. G. Sardegna and A. Jarosz (eds.), *Theoretical and Practical Developments in English Speech Assessment, Research, and Training*, Second Language Learning and Teaching,
https://doi.org/10.1007/978-3-030-98218-8_10

1 Introduction

To have a successful communication in English, L2 learners need to say words that others can understand (i.e., be intelligible) as well as understand the words spoken to them (i.e., have high comprehensibility). In other words, the pronunciation of vocabulary is central to intelligibility (Levis, 2018). To be able to pronounce lexical words clearly, first and foremost, learners of English need to understand the distinction between stressed and unstressed syllables. Stressed syllables tend to be longer, louder, and higher pitched than unstressed syllables (Derwing & Munro, 2015). Different placement of English stress can result in difference in meaning or part of speech (*ímport* vs. *impórt*) (Liu, 2017). Also, when the main stress is not assigned to the correct syllable, the rhythm is distorted and may obscure the meaning of the word (Benrabah, 1997; Field, 2005) and even affect listeners' ability to recognize the word (Cutler, 2012, 2015; Cutler et al., 1997). Moreover, English word stress is not fixed to a given position like in other languages (e.g., Polish stress is fixed to the penultimate syllable), and this difference often causes L1 transfer errors with word stress (Archibald, 1998; Liu, 2017). For these reasons, it has been argued that learners of English would benefit from instruction on how to stress polysyllabic words (Levis, 2018).

A growing number of studies provide evidence that explicit pronunciation instruction works (Lee et al., 2015; Sardegna & McGregor, 2022; Thomson & Derwing, 2015). Yet, complaints about time constraints to incorporate pronunciation instruction in language classrooms are abundant (Foote et al., 2011), which causes teachers not to devote class-time for teaching and practicing English word stress.

This chapter explores an alternative to pronunciation classroom instruction for teaching English word stress. After a brief review of prior literature on approaches to teaching English word stress, it reports on a study that explored students' autonomous learning efforts for improving their ability to predict and perceive English word stress through the use of worksheets and *YouGlish* (www.youglish.com)—a free *YouTube*-based pronunciation dictionary. The chapter concludes with a discussion of the implications of the findings and areas for future research.

2 Literature Review

There has been an increasing number of studies investigating pronunciation classroom research with a focus on the effectiveness of explicit instruction alone or combined with other instructional components, such as awareness-raising, perceptual training, oral production practice, and corrective feedback (for a systematic review, see Sardegna & McGregor, 2022). Most research on awareness-raising has shown the positive effects of prosodic feature awareness training through explicit instruction of rules and strategies (Luchini, 2017; Sardegna, 2009, 2012, 2021).

Recent research on perceptual training has combined instruction and awareness-raising through typographical enhancements followed by extended listening input. For example, Kartal and Korucu-Kis (2020) provided explicit instruction for target words using *Twitter* followed by links for listening to the target via *YouGlish* as a follow-up activity, and found that these two resources supported students in learning and retaining commonly mispronounced words. The effects of explicit instruction followed by oral production practice through focus on form have received the most attention (see Thomson & Derwing, 2015). Additionally, based on her prior research on individual learner variables, Sardegna (2021) has argued for the need to offer guided focus-on-form practice and combine it with other instructional components that can enhance the learning process: goal prioritization, awareness raising, ongoing feedback, and opportunities for reflection on progress. These components rely on the critical role of the teacher in supporting the learning process. Yet, given that many foreign language (FL) classrooms do not have much class time to devote to this entire process, the question remains as to which of these components are the most effective in helping students improve their pronunciation. Also, technology can easily bring authentic materials and different speech accents into the FL classroom and serve as good models for practice (Sardegna & Hughes, 2022). Investigating ways of incorporating instructional technology resources in support of pronunciation instruction may offer a solution to the time limitations of the classroom teacher.

Furthermore, many studies have shown evidence suggesting the positive benefits of explicit pronunciation instruction on suprasegmentals and the pronunciation of academic words (Derwing et al., 1998; Kartal & Korucu-Kis, 2020; Sardegna, 2012). Comparatively fewer have assessed progress with English word stress alone. Tanner and Landon (2009) found evidence suggesting that self-directed computer-assisted practice helped 75 ESL students improve their perception and production of English word stress. Sardegna (2009, 2021) reported findings with respect to the production of English word stress from two semester-long interventions involving explicit pronunciation instruction, awareness-raising activities, and frequent and guided practice on form using pronunciation learning strategies. In both studies, the ESL learners significantly improved their ability to stress polysyllabic words after they received explicit pronunciation instruction in a pronunciation course (4 months). In addition, the students maintained significant long-term progress, and in Sardegna (2021) the experimental group significantly outperformed a control group. These studies provide strong evidence that English word-stress rules are teachable and learnable. The current study sought to investigate whether autonomous learning of English word-stress rules (i.e., without the help of the teacher) would produce similar results.

3 The Study

This study explored high-school EFL students' autonomous learning gains after the teacher gave them four worksheets with information about word-stress rules and encouraged them to practice the rules out-of-class and with the help of *YouGlish*. The research questions were the following:

1. To what extent did the EFL students improve their ability to predict and perceive English word stress after practicing two word-stress rules autonomously for four weeks?
2. What was learners' practice engagement and views on the usefulness of word-stress rules and *YouGlish* for improving English word stress in polysyllabic words?
3. Did learners' motivations and practice engagement contribute to differences in learning outcomes with respect to predicting and/or perceiving English word stress?

4 Method

4.1 Participants

The participants were 12 (9 male and 3 female) Polish EFL learners taking a language course at a state secondary school in Poland. They were 16–18 years old and at an upper-intermediate level of English. They had started learning English in kindergarten at the age of 5–6. They had no prior knowledge of English word-stress rules and, when recruited for the study, they all enthusiastically volunteered to participate acknowledging a certain struggle to stress English polysyllabic words and hoping that the information and practice would be helpful. English word-stress rules are not taught in language classes because this topic goes beyond the scope of the national curriculum and these students' oral school-leaving exam requirements.

4.2 Learning Materials

Four instructional worksheets based on Dickerson's (2004) word-stress prediction system were developed for students' out-of-class learning and practice. Dickerson's (2004) system consists of four orthographic word-stress rules. The rules include information on prefixes, suffixes, and vowel and consonant combinations, and direct learners' attention to two main syllables in the word: the key syllable (KEY) and the left syllable (LEFT). The KEY can be found immediately left of an ending and the LEFT can be found immediately left of the KEY, as in the following example:

> e.g., "approximate" → appr*óx*<u>im</u>(ate, whereas -ate is the ending, the underlined syllable is the KEY, and the italicized syllable is the LEFT.

Learners can predict the major stress of a word by identifying the word's ending, the KEY and LEFT, and then applying the word-stress rule that corresponds to the word category (noun, adjective, verb) and the identified ending (Dickerson, 2004, 2015). Polysyllabic words are stressed either on the KEY or the LEFT so it is important that these two syllables are identified correctly. For example, "approximate" is stressed on the LEFT.

Participants focused on two of the four word-stress rules proposed by Dickerson (2004): the Key Stress Rule (KSR) and the Left Stress Rule (LSR). KSR directs learners to stress the KEY and applies to words ending in *-ia, -io, -iu, -ienC* (C = Consonant). These iV (V = Vowel) endings can be followed by a consonant or other endings (e.g., *-ed*, *-ing*), including endings for other rules (e.g., *-ive, -able, -ated*). That is, in making stress predictions, iV endings should be considered before other endings. In the following examples, the KEY is underlined, iV endings are in bold and marked with an open parenthesis, and other endings are marked with a square bracket:

e.g., méd(**ia**, nutrít(**io**us, musíc(**ia**n theoretíc(**ia**n, obéd(**ienc**[e, inít(**iat**[ive, fásh(**io**n[able, appréc(**iat**[ive, repúd(**iat**[ing, opín(**io**n[ated

LSR directs learners to stress the LEFT and applies to long words (of three or more syllables) ending in -*y, -ate, -ated, -ator, -ating, -acy, -acies*. In the following examples, the LEFT is in italics and the KEY is underlined:

e.g., appr*óx*im(ate, prox*ím*it(y, *ánn*ot(ated, s*ímpl*if(y, int*ím*id(ating, regul*ár*it(y, fac*íl*it(ate pr*ócr*e(ating, indet*érm*in(acy

Also, this system allows learners to make sound-spelling predictions (Dickerson, 2012). After assigning the major stress to the appropriate syllable, learners can predict stressed and unstressed vowels and vowels left of the major stress by applying orthographic vowel quality patterns (see Dickerson, 2004, 2012, 2015). Knowing which vowel to produce is as important as knowing which syllable to stress because native speakers of English pay attention to the quality of the vowel in order to determine if a syllable is stressed (Cutler, 2015).

To scaffold participants' learning, a new worksheet was provided every week. Worksheet #1 (Week 1) focused on identifying endings, and the KEY and LEFT. Worksheets #2 and #3 (Weeks 2–3) offered information and practice on KSR and LSR words, respectively, including how to identify the syllable to stress and how to pronounce the stressed vowel. Worksheet #4 (Week 4) provided additional practice with the two rules (see Dickerson, 2004, 2015; Hahn & Dickerson, 1999, for more information). All the worksheets included answer keys for self-correction. Participants were asked to work on these worksheets autonomously out of class and at their own pace for four weeks (no teacher feedback was provided). Participants were also told that their improvement with word stress would be assessed at the end of the four weeks.

For perception practice, participants were encouraged to use *YouGlish* (www.youglish.com). *YouGlish* is a free YouTube-based site that has more than 100 million tracks of speakers of different varieties of English pronouncing words in context. Users can search for a word and automatically get short video clips showing how native people use that word in a real context (e.g., a speech, an academic presentation, an interview). Participants could use this resource for extensive listening as *YouGlish* enables them to listen to one video clip after another by clicking on an arrow, and to pause and listen to each video multiple times.

4.3 Data Collection and Analysis

There were four sources of data collection: a background questionnaire, a pronunciation tracker, a prediction test taken twice, and a perception test taken twice. The background questionnaire elicited information about participants' characteristics (gender, age, years of studying English), knowledge of word-stress rules, self-assessed pronunciation weaknesses and strengths, views on using online resources in pronunciation practice, and motivations for pronunciation learning. The pronunciation tracker was used to gather information about what, how, when, and how much participants practiced during the four weeks of the study, and their opinions on the usefulness of their practice choices.

The students took the prediction test followed by the perception test twice (pre- and post-instruction). Both tests consisted of the same list of 40 polysyllabic words, which included 10 words stressed by KSR, 10 words stressed by LSR, and 20 words stressed by two other word-stress rules (10 each). Hence, the list had as many KSR and LSR words as distractors. Each word had four choices, each with a different syllable capitalized to indicate the placement of the major stress in the word. For the prediction test, participants could take as much time as they needed to figure out which of the four choices corresponded to how the word was stressed. They could not consult any resources during the test. For the perception test, participants listened to a recording of four different native speakers of English taking turns at reading the words aloud. The speakers read each word twice using a falling intonation both times. There was a period of silence between target words for participants to read the four choices and select the one that corresponded to how they "heard" the speaker pronounce/stress any given target word.

To determine improvement with ability to predict and perceive English word stress, prediction and perception pre- and post-tests scores were compared using Wilcoxon signed-rank tests due to the small sample size and because the distribution of the scores by stress rule was not normal (Field, 2013). Word-stress production changes (including changes in vowel quality) were not assessed although the students were also encouraged to practice reading the words aloud. Following Rosenthal (1991), the effect size of the obtained results (practical significance) was calculated by dividing the z value by the square root of the number of observations, whereas $r > 0.50$ = large effect. To ascertain whether learners' practice engagement and views mattered, two groups of participants (high engagement and low engagement groups) were formed based on participants' entry responses on their pronunciation trackers. To examine participants' overall opinions regarding the usefulness of the two word-stress rules and *YouGlish*, a thematic analysis of their responses to the question "What is your opinion of the rules/*YouGlish*?" was performed by group.

The participants were prompted to answer this question every time they entered an activity in their pronunciation trackers.

5 Results

5.1 Learners' Characteristics

The background questionnaire revealed that none of the participants had studied English pronunciation before or had prior knowledge of English word-stress rules. When asked about their English pronunciation strengths, half responded not knowing which they were. Those that were able to identify their strengths mentioned speaking loudly, clearly, and fluently; learning quickly; learning from watching movies in English and listening to English music; and learning from watching sports and series in English. Interestingly, when asked about their pronunciation challenges, only one student identified a sound problem (with the consonant cluster /rl/). All others mentioned problems at the word level: eight said they struggled with pronouncing new and long/difficult English words, and three mentioned problems with understanding such words when uttered by native speakers. The following comments illustrate their pronunciation concerns:

- *My challenges are difficult English words.*
- *New words, especially the long and complicated ones.*
- *Mostly new words I never heard before.*
- *My problem is repeating same words while saying something to someone. After that I miss a point and barely can say something, I am getting nervous.*
- *My weak point is that I think too much before saying a word that I'm not sure about.*
- *Pronunciation of words without wrongs, like native speakers.*
- *Sometimes I don't know relevant word and I don't know how to continue speaking.*
- *Sometimes it is hard for me when I must read a word which is new for me.*
- *It's hard for me to understand when interlocutor is speaking really fast.*
- *I have a hard time understanding words pronounced quickly in English.*
- *It can be hard for me to understand when someone is fast speaking, or sometimes when I am listening the music. For example, I hear other words or I have an impression of hearing totally new language.*

When asked what they wanted to improve regarding their English pronunciation, three mentioned their accent, with four others also expressing a desire to speak fast, fluently and/or confidently in English (e.g., "*I want to speak more fluently and confidently, without a thick Polish accent.*"). In addition, two participants mentioned their need to better understand English lyrics and words pronounced quickly in English. Two others aspired to have correct English intonation for words or improve the pronunciation of words. In sum, most participants seemed most concerned with how to understand and produce English (long) words. While they did not directly mention English word stress as a concern or goal for learning, instruction on English word stress can address both their production and comprehension needs as it has been shown that misplacement of English word stress can affect how comprehensible a speaker is perceived to be (Anderson-Hsieh et al., 1992), and English word stress can affect spoken word recognition (Cutler, 2012, 2015).

Furthermore, participants reported no to little knowledge (and no use) of apps or online resources for pronunciation learning, but all unanimously agreed that online resources could be very useful and helpful for improving their pronunciation skills. As the following responses show, they appreciated the fact that online resources could help them access native speaker models and allow them to practice on their own, any time, for free, and as much and as frequently as needed/desired:

Online resources are (very) helpful because…

- *we can improve our pronunciation on our own, without teacher.*
- *they are easy to use and accessible.*
- *many of them are free and … the access to them is really easy. Apart from that we can study anywhere and when we want.*
- *we can use it anytime and anywhere we want and usually resources like apps that recognize our speech can verify correctness of intonation and other aspects of pronunciation.*
- *are useful when you practice them regularly.*
- *we can learn from native speaker.*
- *we can watch movies in English or listen to the broadcast.*
- *you can practise your English in home and you can learn a lot of new things.*
- *every type of practising can help, if we are motivated.*
- *in our's times every online source of learning stuff is good for students. They spend a lot of time on Internet.*
- *there are many pages that show how to pronounce words correctly.*

Overall, participants (six agreed, six strongly agreed) reported high motivation to improve their pronunciation, and most (two neutral, seven agreed, three strongly agreed) expressed they would like to use online resources to improve their pronunciation skills on their own. Also, as revealed by their comments, most acknowledged that they needed and wanted help with producing and understanding English words. Hence, this seemed an ideal group of students to test the efficacy of guided autonomous pronunciation learning based on students' stated needs and wants.

5.2 Autonomous Learning Outcomes

Pairwise T1-T2 comparisons using Wilcoxon signed-rank tests were computed to explore differences in improvement with respect to participants' stress predictions and perceptions for words stressed by KSR and LSR after their four weeks of autonomous learning. The tests revealed no significant differences between T1 and T2 mean scores for any of the rules (KSR and LSR) or skills (prediction and production) (see Table 1). This finding suggests that, despite participants' high motivations to improve their pronunciation skills, their autonomous learning behaviors were not successful enough to effect changes of statistical significance in their abilities to predict and perceive word stress. Figure 1 displays participants' scores across time.

As observed in Fig. 1, participants' prediction and perception scores for KSR words were higher at T2; yet this difference did not reach statistical significance. The lack of statistical significance may be due to high variability in a small sample (see Table 1). As participants were left to decide on their own what to do, how often,

Table 1 Means, standard deviations, medians, and results of related-samples Wilcoxon signed-rank tests for prediction and perception scores by word-stress rule ($N = 12$)

Stress rule	*M* (*SD*), *Mdn*	*T*	*z*	*p*	Effect size (*r*)
KSR prediction	T1: 39.17 (18.81), 35.00 T2: 61.67 (33.53), 75.000	54	1.873	0.061	0.38
KSR perception	T1: 34.17 (18.32), 35.00 T2: 51.67 (33.26), 55.00	61.5	1.780	0.075	0.36
LSR prediction	T1: 55.00 (21.53), 55.00 T2: 56.67 (26.74), 55.00	42	0.244	0.807	0.05
LSR perception	T1: 60.00 (17.06), 60.00 T2: 60.00 (23.74), 60.00	14	0.000	1	0.00

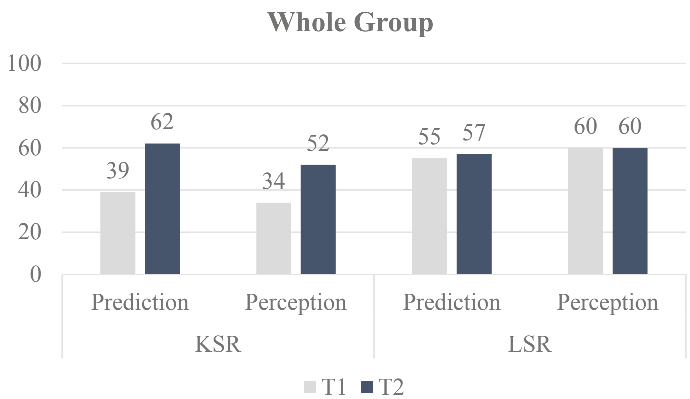

Fig. 1 Prediction and perception scores by rule (KSR and LSR) and time (T1 and T2) for the whole group

and how much, it is possible that their behavioral choices and views on how best to practice word stress affected their progress both positively and negatively, thereby creating a higher variability at T2 (i.e., some improving and some decreasing in their predictions and perceptions). To determine if there were any differences in participants' autonomous behaviors that could explain the group findings, we examined participants' pronunciation trackers for evidence of their practice engagement and views.

5.3 Learners' Practice Engagement and Views

After confirming participants' high motivations to improve their pronunciation on their own using online resources at the beginning of the study, it was important to determine their actual practices during the four weeks of the study. Based on participants' pronunciation trackers documenting their daily practice engagement with word-stress rules and *YouGlish*, we divided the class evenly in two groups: the high engagement (HE) group and the low engagement (LE) group. Konrad, Darek, Marek, Arek, Alina, and Nikolai constituted the HE group because they reported completing *all* the worksheets (1–4) and using *YouGlish* for perception practice in the manner indicated by the instructor. In contrast, Borys, Pavel, Julia, Albin, Zuzana and Aleksy formed the LE group because they reported no to little practice with the worksheets, and some did not use *YouGlish* for perception practice at all. Table 2 shows participants' accumulated practice time, number of practice entries, and materials used for practice by group (names are pseudonyms).

A comparison of both groups at T1 using Mann–Whitney U-tests revealed no significant differences between the two groups at T1 for (a) predicting KSR words (HE $Mdn = 35$, LE $Mdn = 40$, $U = 18.5$, $z = 0.081$, $p = 1.000$, $r = 0.02$); (b) perceiving KSR words (HE $Mdn = 30.00$, LE $Mdn = 35.00$, $U = 21.5$, $z = 0.574$, $p = 0.589$, $r = 0.12$); (c) predicting LSR words (HE $Mdn = 55.00$, LE $Mdn = 55.00$, $U = 15.5$, $z = -0.407$, $p = 0.699$, $r = -0.08$); and (d) perceiving LSR words (HE $Mdn = 60.00$, LE $Mdn = 50.00$, $U = 13$, $z = -0.815$, $p = 0.485$, $r = -0.17$). These results suggested the comparability of the two groups at the beginning of the study and, as already discussed, the groups were also comparable with respect to their high motivations to improve their English pronunciation skills.

A thematic analysis of participants' comments in their pronunciation trackers further revealed that the HE group found the word-stress rules useful, interesting, easy to learn, and helpful. They also reported using *YouGlish* and finding it useful, helpful, easy to use, and an interesting and accessible form of learning. Some sample comments follow:

- *I would recommend it* [worksheet # 1] *for everyone because it was fast, nice and easy to remember.* (Arek)
- *It helped me with new words. The simplest part was reading and the hardest part was listening.* (Alina)

Table 2 Practice time and materials used by group

Group	Name	Total practice time in hours	*N* of entries	Materials used
HE[a]	Konrad	5:00	4	All worksheets + *YouGlish*
	Darek	3:10	5	All worksheets + *YouGlish*
	Marek	3:00	5	All worksheets + *YouGlish*
	Arek	2:30	5	All worksheets + *YouGlish*
	Alina	2:00	6	All worksheets + *YouGlish*
	Nikolai	2:00	4	All worksheets + *YouGlish*
LE	Borys	1:00	1	*YouGlish*
	Pavel	0:40	2	Worksheet #2 (KSR) + *YouGlish*
	Julia	1:15	3	Worksheet #2 (KSR) + *YouGlish*
	Albin	1:10	4	Only read the worksheets (no practice)
	Zuzana	2:00	7	Worksheets #2 (KSR) and #3 (LSR)
	Aleksy	3:00	6	*Youglish*, Worskheet #4 (review) for one hour on the day before the test

[a]HE = High Engagement group; LE = Low Engagement group

- *This pdf* [worksheet #2] *is well made and it is nice to read and easy to understand all of the rules. Everyone should try these exercises. It is really helpful and quite easy if someone wants to learn and is doing it systematically.* (Arek)
- *Interesting and helpful. Learned a few things.* (Nikolai)

Furthermore, Darek, Marek, and Konrad practiced for 3–5 h overall and reported being highly motivated to continue the practice on their own:

- *I would recommend this activity to others, because I think that watching educational films on YouTube is an interesting and accessible form of learning.* (Darek)
- *I can see improvement since we started learning about the stress rules. Now I can recognise the LSR and KSR in words. While I was listening to the words I could better hear the stress.* (Darek)
- *After doing all these exercises and practicing pronunciations with YouGlish, I know and understand much more and I'm sure it'd help even more if I had spent more time with it. And that's what I'm up to:).* (Marek)
- *I think that I did better… because I could apply rules. I feel like I haven't improved much in speaking, because I wasn't practising it too much, so now I plan to focus more on speaking exercises.* (Konrad)

In contrast, as Table 2 shows, the LE group was not as enthusiastic with the stress rules, and thus made different choices for their autonomous practice. For example, Borys did not practice with the worksheets at all. In a comment on his pronunciation tracker, he explained that he disliked going through the information in the worksheets, so he decided to practice for one hour just with *YouGlish,* which he saw as more useful and practical:

- *Cards with exercises (key stress rule) are not giving any chance for progress in my case. There is too much information in one place. YouGlish, in contrast, is very useful in everyday and occasional use.* (Borys)

Three other participants in the LE group also reported little practice with the worksheets. Pavel and Julia only practiced with worksheet #2 (KSR), and Albin reported reading but not doing the practice exercises on the worksheets. Pavel and Alvin acknowledged some struggle with independent learning and the fact that the information was provided in the L2:

- *Independent learning is not effective enough and there is a need to work in class together with the teacher. Learning direct rules can be confusing and difficult.* (Pavel)
- *I think it would be a little bit easier when the most important rules were in Polish. I think I need more practice with these rules.* (Alvin)

Julia seemed to like the KSR rule, as shown in the following comment, but she also struggled with independent learning because she stopped doing the worksheets after the second week:

- *I would recommend this activity to others, because it is very useful when you want to pronounce a word that you didn't know before.* (Julia)

In addition, despite reporting 6–7 practice activities during the four weeks, which at first sight might seem to indicate high involvement with the materials, Zuzana and Aleksy were also included in the LE group because they did not complete the activities as instructed (see Table 2). Zuzana only worked with two of the four worksheets, each time for about 10/15 min (not enough time to read the information and complete the activities) and did not use *YouGlish* at all. Aleksy mostly engaged with *YouGlish* with the exception of the day before the post-test, in which he reported practicing using worksheet #4 (the review worksheet) for an hour. This worksheet does not explain the rules. It has lists of words for students to apply the new knowledge gained through worksheets #1–3, and an answer key to self-correct. Also, despite their observed attempts at trying to understand the rules, Zuzana and Aleksy reported finding them difficult to understand without the help of the teacher:

- *I work better when someone explains it … I'd rather work more with the teacher.* (Zuzana)
- *Useful* [referring to the worksheets]*, but sometimes it's hard to understand without any help.* (Aleksy)

Table 3 summarizes participants' practice choices and opinions regarding the instructional resources by group.

Table 3 Practice choices and opinions regarding the instructional resources by group

HE[a] ($N = 6$)	LE ($N = 6$)
• Found the stress rules useful, interesting, easy to learn, and helpful	• Found the rules not effective/useful OR found the rules useful yet hard to understand
• Found *YouGlish* helpful, easy to use, useful, interesting, and an accessible form of learning	• Found *YouGlish* helpful (made learning easier), interesting/useful, and the best method to learn English. [Note: 2 reported not using *YouGlish*.]
• Wanted to continue practicing because they could see improvements in their phonological awareness and listening comprehension	• Most gave up practice after learning about KSR
• Recommended teaching the rules for self-practice	• Expressed preference for working with a teacher rather than alone

[a]HE = High Engagement group; LE = Low Engagement group

Table 4 Improvement with word-stress predictions and perceptions by rule (KSR and LSR) and group (HE and LE)

Group	Stress rule	*M* (*SD*), *Mdn*	*T*	*z*	*p*	*r*
HE	KSR prediction	T1: 38.33 (20.41), 35.00 T2: 88.33 (4.082), 90.00	21	2.207	0.027*	0.64
	KSR perception	T1: 31.67 (23.17), 30.00 T2: 75.00 (20.74), 70.00	21	2.214	0.027*	0.64
	LSR prediction	T1: 58.33 (19.41), 55.00 T2: 71.67 (29.94), 75.00	16.5	1.294	0.196	0.37
	LSR perception	T1: 63.33 (15.06), 60.00 T2: 71.67 (17.22), 70.00	6	1.633	0.102	0.47
LE	KSR prediction	T1: 40.00 (18.97), 40.00 T2: 35.00 (27.39), 35.00	5	-0.680	0.496	-0.19
	KSR perception	T1: 36.67 (13.66), 35.00 T2: 28.33 (26.40), 30.00	6.5	-0.850	0.395	-0.25
	LSR prediction	T1: 51.67 (24.83), 55.00 T2: 41.67 (11.70), 40.00	4.5	-1.289	0.194	-0.37
	LSR perception	T1: 56.67 (19.66), 50.00 T2: 48.33 (24.83), 55.00	1.5	-1.289	0.197	-0.37

*$p < 0.05$. There was a significant difference between the groups

5.4 Learning Outcomes by Group

Pairwise T1-T2 comparisons using Wilcoxon signed-rank tests were computed to explore differences in improvement with respect to participants' stress predictions and perceptions for words stressed by KSR and LSR by group (HE and LE) after their four weeks of autonomous learning. The test results are displayed in Table 4.

The HE group made significant improvements with predicting and perceiving KSR words and both represented large effects ($r = 0.64$). Also, while their improvements with LSR words were not significant, they exhibited an upward trend of moderate practical significance (Fig. 2). In contrast, the LE group did not make any significant changes. In fact, contrary to expectations, these participants' performance with words stressed by KSR and LSR was worse than at the beginning of the study for both prediction and perception (Fig. 3). These two groups' opposite trends may explain the non-significant findings for the group as a whole.

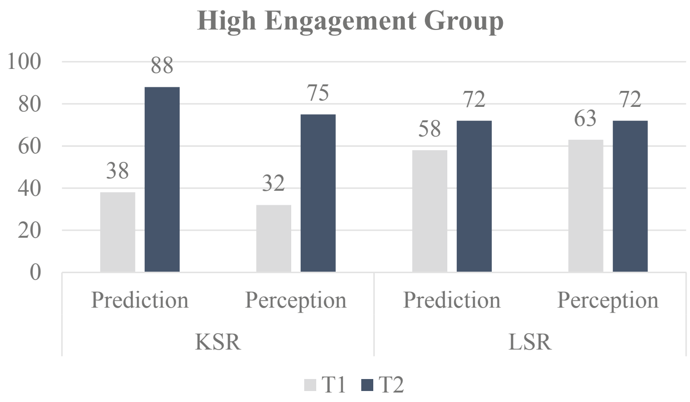

Fig. 2 Prediction and perception scores by rule (KSR and LSR) and time (T1 and T2) for HE

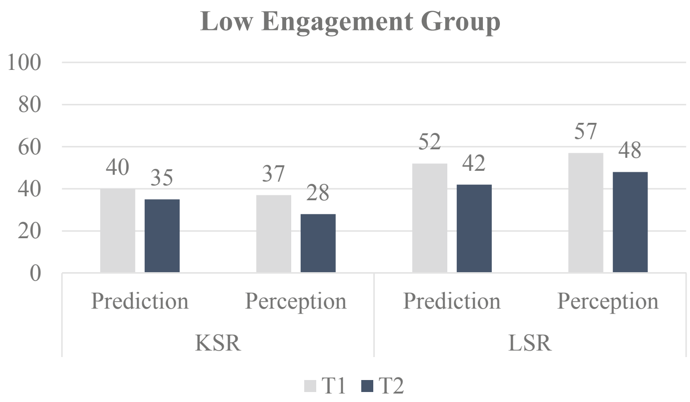

Fig. 3 Prediction and perception scores by rule (KSR and LSR) and time (T1 and T2) for LE

6 Discussion

The results of this investigation confirm previous findings suggesting that word-stress rules are teachable and learnable (Sardegna, 2009, 2012, 2021). The HE group, who reported sustained and self-directed focus-on-form practice via worksheets and extensive listening input via *YouGlish,* was able to make improvements in their word-stress predictions and perceptions. To guide participants' autonomous learning, the worksheets provided explicit instruction and scaffolded practice with two word-stress rules (KSR and LSR) and *YouGlish* offered exposure to native speaker talk, which participants could use for self-correction and as models to imitate. The HE group practiced with the materials as instructed, and provided evidence suggesting that the rules can effectively support autonomous learning of English word stress when students (a) are intrinsically motivated to learn, (b) practice on their own, and (c) see the value of their practice efforts. In contrast, the LE group, who reported little or no engagement in self-directed focus-on-form practice with the worksheets, and either did not use *YouGlish* or used it without learning the rules first (i.e., without increased phonological awareness), did not improve their ability to predict and perceive word stress. These findings underscore the relevance and need of combining three instructional components for pronunciation learning: explicit instruction, increased pronunciation awareness, and focus on form (Sardegna, 2021; Sardegna & McGregor, 2022). Progress in pronunciation largely depends on students' practice efforts. The efficacy of these efforts relies on students' increased declarative knowledge of what they need to improve and how. Without this knowledge and increased phonological awareness of the target feature, their efforts (e.g., working with *YouGlish* without knowing what to listen for) may be in vain, as suggested by the learning outcomes of the LE students who reported working only with *YouGlish.*

The HE group found the rules useful and worth their time and effort and, consequently, made considerable improvements in their recognition and prediction of English word stress. They also found it useful and helpful to spend time with extensive listening input via *YouGlish*, which, as they reported, helped increase their awareness and confidence in their knowledge of English word stress. The findings also indicated that it was easier for the HE group to improve KSR words (significant large effects) than LSR words (positive trend upwards but non-significant findings). It is possible that LSR words are more difficult because they require more processing time as learners need to memorize more endings than for KSR words. It is also possible that these learners noticed more striking differences in their predictions/perceptions of KSR words given their lower initial levels (38/32% for KSR words compared to 58/63% for LSR words; see Table 4) and thus focused more on improving their stress predictions and perceptions of KSR words. Also, larger improvements with KSR words may just be related to the fact that these learners had more room to grow than with LSR words. All in all, the HE group found the resources useful, seemed to enjoy their autonomous learning activities, and effected changes with word-stress predictions and perceptions in the right direction. Future research might want to investigate other ways of providing support to maximize highly engaged students' practice efforts and outcomes.

However, despite the positive outcomes with the HE group, we cannot disregard the fact that they constituted half of the class. At the beginning of the study, participants in the LE group reported to be as eager as the HE group to learn English pronunciation and use online resources autonomously; yet, they struggled with understanding and using the stress rules and repeatedly voiced the need to have teacher support. As a result, they either gave up on the worksheets after a few attempts or simply resorted to just working with *YouGlish*. Another interesting finding is that the LE group performed worse, which suggests two plausible interpretations: (a) they were choosing stress assignments at random, or (b) once they became aware of their tendency to make wrong word-stress predictions/perceptions (remember that most worked with worksheet #1 and then gave up), they must have second-guessed their stress predictions/perceptions at T2 in an effort to do better, which resulted in an increase in errors since they never took the time to learn the rules.

A further factor that may have affected the outcomes is that students' behavioral practices completely relied on their intrinsic motivations and willingness to improve, as their work was not assessed or graded by the teacher—that is, students' (lack of) work/improvement bore no consequences in their academic studies. Notoriously, the whole group clearly expressed a high motivation to work on their pronunciation skills at the beginning of the study; yet, half of the group lost interest soon after the study started. This observation reinforces the view that motivation is a dynamic construct. As argued by Beltman and Volet (2007), sustained motivation is inextricably linked to both the person and context, mediated by individuals' ongoing appraisal process (e.g., high/low enjoyment), and constantly revised as a result of changes in personal and contextual circumstances (e.g., loss of interest, lack of support). Comments from the HE group showed that their high positive appraisals regarding the learning experience (e.g., "*I can see improvement since we started learning about the stress rules*"—Darek; "*I know and understand much more*"—Marek; "*I think I did better*"—Konrad) led them to persist in practicing. As the study progressed, their confidence regarding the rules (what to do and how) increased and they experienced success in their efforts, which may have increased their sense of self-efficacy, and consequently, their willingness to continue practicing. This finding supports Sardegna et al.'s (2018) model, which demonstrated that the higher students' self-efficacy beliefs are, the more likely they are to find ways, time, and strategies to improve their pronunciation. In contrast, as the study progressed, the LE group became increasingly more confused and worried (e.g., *"Learning direct rules can be confusing and difficult"*—Pavel; *"Cards with exercises are not giving any chance for progress in my case"*—Borys). Their negative appraisals of the learning experience led them to discontinue involvement (as argued by Beltman & Volet, 2007). Hence, as shown by the study findings and in Sardegna et al. (2018), students' low sense of accomplishment may critically affect their learning motivation and behavioral intentions. Future research might want to investigate if providing some kind of extrinsic motivation to students with low self-efficacy/success may help nurture their intrinsic motivations and increase their feelings of success/accomplishment so that they are more likely to persist in their autonomous practice efforts when they face challenging tasks. Such extrinsic motivation may be in the form of teacher feedback (Sardegna, 2021), grades, and guided self-reflections on learning gains (Sardegna & McGregor, 2013, 2017). Arguably, teacher support in terms of specific assignments and ongoing feedback (i.e., a classroom structure) could have increased the motivations of the LE group to practice the rules as well as their understanding of the rules, and ultimately, help them improve, too. While there is increasing evidence suggesting the important role of the teacher in providing explicit pronunciation instruction, feedback, and support through coursework and individualized student–teacher meetings (Sardegna, 2009, 2012, 2021), so far little is known about how the teacher can support and help sustain the motivations of *all* the students in pronunciation tasks completed alone and out-of-class without also requiring extensive individualized and ongoing teacher feedback—namely, what we hoped to avoid in this study. This area of research warrants further exploration.

Finally, the results offer evidence in favor of *YouGlish's* potential for educational purposes, which echo findings from Kartal and Korucu-Kis (2020). The ten participants that used *YouGlish* found it a useful tool for autonomous learning. However, only the HE group, who reported using *YouGlish* in the manner instructed—that is, after learning the rules so that they could use their increased awareness of English word stress to listen for the major stress in the target word and practice it—showed improvements in their perceptions in the post-test. Thus, it seems that extensive perceptual practice enhances learning, but it needs to follow raised awareness for it to be most effective in helping students improve their perception of English word stress.

7 Pedagogical and Research Implications

In addition to corroborating prior research on the learnability of English word stress, the findings provided further insights regarding the learnability of two specific word-stress rules: KSR and LSR. They also extended our understanding of two other factors worth considering to maximize learning outcomes for *all* the students:

1. A classroom structure with assigned tasks and a time in class devoted for practice is not essential as the HE group did not need it, but may be necessary to support and engage *all* students in autonomous practice.
2. Teacher feedback may not be essential as the HE group did not need it, but may be necessary to keep *all* students on track, increase their sense of self-efficacy, re-teach concepts as necessary, and provide encouragement through positive reinforcement.

This study has some limitations that should be considered in future research. First, it involved L1 Polish EFL high-school students and a small sample size. Future research might want to extend the findings to other populations and settings.

Second, it only investigated prediction and perception outcomes with two word-stress rules. It would be important to corroborate the findings with other stress rules, and with production scores in read-aloud and free-speech conditions. Third, it did not assess maintenance of improvement over time or the effects of improvement with word stress on L2 comprehensibility and intelligibility. These two areas need attention. Finally, the study offered valuable insights regarding students' needs when learning autonomously without teacher support; yet the findings need to be corroborated with larger groups. Another fruitful area of future research is to investigate ways of enhancing and maintaining the autonomous learning efforts of learners that experience low success in their learning attempts.

8 Conclusions

This study explored an alternative approach to pronunciation classroom instruction. Instead of learning directly from the teacher, twelve high-school EFL students received four worksheets with information about two word-stress rules (KSR and LSR) for out-of-class autonomous learning and were encouraged to use *YouGlish* for extensive listening support for four weeks. Despite their high initial motivations to practice, only half of them (HE group) practiced with all the materials and improved in their confidence and ability to predict and perceive English word stress (large effects with KSR words). The other half (LE group) struggled with the materials and, consequently, discontinued the practice, resulting in negligible learning gains. The findings underscore the need to nurture and enhance students' motivations to learn so that they put the time and effort needed to effect changes in their pronunciation skills. Future research might want to investigate the role of the teacher in supporting *all* students in their autonomous learning efforts. Would adding feedback support be enough to engage *all* learners in out-of-class focus-on-form practice with English word stress, or classroom time and practice would also be needed? This question merits further investigation.

References

Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgements of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning, 42*(4), 529–555. https://doi.org/10.1111/j.1467-1770.1992.tb01043.x

Archibald, J. (1998). *Second language phonology.* John Benjamins.

Beltman, S., & Volet, S. (2007). Exploring the complex and dynamic nature of sustained motivation. *European Psychologist, 12*(4), 314–323. https://doi.org/10.1027/1016-9040.12.4.314

Benrabah, M. (1997). Word stress: A source of unintelligibility in English. *IRAL, 35*(3), 157–165. https://doi.org/10.1515/iral.1997.35.3.157

Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words.* MIT Press.
Cutler, A. (2015). Lexical stress in English pronunciation. In M. Reed & J. M. Levis (Eds.), *The handbook of English pronunciation* (pp. 106–124). Wiley Blackwell. https://doi.org/10.1002/9781118346952
Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech*, *40*(2), 141–201. https://doi.org/10.1177/002383099704000203
Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research.* John Benjamins. https://doi.org/10.1075/lllt.42
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
Dickerson, W. B. (2004). *Stress in the speech stream: The rhythm of spoken English.* University of Illinois Press.
Dickerson, W. B. (2012). Prediction in teaching pronunciation. In C.A. Chapelle (Ed.), *The encyclopedia of applied linguistics.* Blackwell. https://doi.org/10.1002/9781405198431.wbeal0950
Dickerson, W. B. (2015). Using orthography to teach pronunciation. In M. Reed & J. Levis (Eds.), *The handbook of English pronunciation* (pp. 488–504). Wiley Blackwell. https://doi.org/10.1002/9781118346952.ch27
Field, A. (2013). *Discovering statistics using IBM SPSS Statistics* (4th ed.). Sage.
Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly, 39*(3), 399–423. https://doi.org/10.2307/3588487
Foote, J., Holtby, A. K., & Derwing, T. M. (2011). Survey of the teaching of pronunciation in adult ESL programs in Canada, 2010. *TESL Canada Journal*, *29*(1), 1–22. https://doi.org/10.18806/tesl.v29i1.1086
Hahn, L., & Dickerson, W. B. (1999). *Speechcraft: Workbook for international TA discourse.* University of Michigan Press.
Kartal, G., & Korucu-Kis, S. (2020). The use of Twitter and YouGlish for the learning and retention of commonly mispronounced English words. *Education and Information Technologies, 25*, 193–221. https://doi.org/10.1007/s10639-019-09970-8
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. M. (2018). *Intelligibility, oral communication, and the teaching of pronunciation.* Cambridge University Press. https://doi.org/10.1017/9781108241564
Liu, D. (2017). The acquisition of English word stress by Mandarin EFL learners. *English Language Teaching, 10*(12), 196–201. https://doi.org/10.5539/elt.v10n12p196
Luchini, P. (2017). Measurements for accentedness, pause and nuclear stress placement in the EFL context. *Ilha do Desterro*, *70*(3), 185–200. https://www.scielo.br/j/ides/a/DyJzKmZRHptFCQvNPYw8h7K/?lang=en
Rosenthal, R. (1991). *Meta-analytic procedures for social research* (Vol. 6). Sage. https://doi.org/10.4135/9781412984997
Sardegna, V. G. (2009). *Improving English stress through pronunciation learning strategies* (Publication No. 3363085) [Doctoral dissertation, University of Illinois at Urbana-Champaign]. ProQuest Dissertations Publishing.
Sardegna, V. G. (2012). Learner differences in strategy use, self-efficacy beliefs, and pronunciation improvement. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 3rd Pronunciation in Second Language Learning and Teaching Conference,* ISSN 2380–9566, Ames, IA, September 16–17, 2011 (pp. 39–53). Iowa State University.

Sardegna, V. G. (2021). Evidence in favor of a strategy-based model for English pronunciation instruction. *Language Teaching*. 1–16. Advance online publication. https://doi.org/10.1017/S0261444821000380

Sardegna, V. G., & Hughes, J. E. (2022). Teaching and learning languages with technology. In M. D. Roblyer & J. E. Hughes (Eds.), *Integrating educational technology into teaching* (9th ed, pp. 332–362). Pearson.

Sardegna, V. G., Lee, J., & Kusey, C. (2018). Self-efficacy, attitudes, and choice of strategies for English pronunciation learning. *Language Learning, 68*(1), 83–114. https://doi.org/10.1111/lang.12263

Sardegna, V. G., & McGregor, A. (2013). Scaffolding students' self-regulated efforts for effective pronunciation practice. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference,* ISSN 2380–9566, Vancouver, British Columbia, August 24–25, 2012 (pp. 182–193). Iowa State University.

Sardegna, V. G., & McGregor, A. (2017). Oral communication for international graduate students and teaching assistants. In J. Murphy (Ed.), *Teaching the pronunciation of English: Focus on whole courses* (pp. 130–154). University of Michigan Press.

Sardegna, V. G., & McGregor, A. (2022). Classroom research for pronunciation. In J. M. Levis, T. M. Derwing, & S. Sonsaat-Hegelheimer (Eds.), *Second language pronunciation: Bridging the gap between research and teaching* (pp. 107–128). Wiley.

Tanner, M. W., & Landon, M. M. (2009). The effects of computer-assisted pronunciation readings on ESL learners' use of pausing, stress, intonation, and overall comprehensibility. *Language Learning & Technology, 13*(3), 51–65. https://www.lltjournal.org/item/10125-44191/; http://dx.doi.org/10125/44191

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

Veronica G. Sardegna, Ph.D. is Adjunct Faculty at Duquesne University in Pittsburgh, USA. She has taught ESL and teacher education courses at the University of Illinois at Urbana-Champaign, the University of Texas at Austin, the University of Pittsburgh, and Duquesne University, USA, for two decades. She conducts research on English pronunciation teaching, intercultural learning, and instructional technology. She has received the 2021 D. Scott Enright Interest Section Service Award for her outstanding service to TESOL.

Anna Jarosz is Assistant Professor in the Department of English Language and Applied Linguistics at the University of Lodz in Poland. Her professional interests include pronunciation teaching and learning with a focus on individual learner differences, motivation, and strategy use. She is author of English Pronunciation in L2 Instruction. The Case of Secondary School Learners (Springer, 2019) and, since 2019, she organizes the International Conference on Native and Non-Native Accents ('Accents').

Speech Production

Vowel Accentedness in the Light of Internal and External Competence Assessment

Jan Volín, Tanja Kocjančič Antolík, Radek Skarnitzl, and Pavel Šturm

Abstract It is generally accepted that vowel quality contributes to the overall impression of accentedness of speech, and affects both intelligibility and comprehensibility. The vocalic oppositions maintain the functionality of the language and past research has identified beg × bag contrast as the most troublesome for Czech learners of English. In this study, we add new empirical data on this contrast, but importantly, we also compare it with pot × port contrast, which is to some extent analogical but also different in both phonological and phonetic sense. Moreover, we relate our findings to (a) internal (own) and (b) external (experienced teachers') assessments of Czech EFL learners' pronunciation competence. Speech samples of 38 Czech learners of English show that individuals differ in their manifestations of foreign accent and that both internal and external assessments reflect the feature of openness in front and back vocalic pairs even in general impression assessments (i.e., without any focus on vowels). This finding highlights the importance of the openness feature in teaching EFL to Czech learners. In addition, the study provides specific formant values that can be used as reference data for cross-linguistic comparisons in future research.

Keywords Intelligibility · Comprehensibility · Native English accents · Non-native English accents

J. Volín (✉) · T. K. Antolík · R. Skarnitzl · P. Šturm
Institute of Phonetics, Charles University, Prague, Czech Republic
e-mail: jan.volin@ff.cuni.cz

T. K. Antolík
e-mail: tanja.kocjancicantolik@ff.cuni.cz

R. Skarnitzl
e-mail: radek.skarnitzl@ff.cuni.cz

P. Šturm
e-mail: Pavel.Sturm@ff.cuni.cz

V. G. Sardegna and A. Jarosz (eds.), *Theoretical and Practical Developments in English Speech Assessment, Research, and Training*, Second Language Learning and Teaching,
https://doi.org/10.1007/978-3-030-98218-8_11

1 Introduction

Foreign-accented speech is a phenomenon of continuous interest throughout human history, and the past decades have provided a more rigorous methodological framework of its research. We no longer base our claims on anecdotal observations, and the space for speculative explanations has been restricted quite substantially. Instead, quantitative empirical research is supported with its clear hypothesising and hypothesis testing. Such research has already resulted in detailed categorisations of various aspects of speech that contribute to accentedness.

One of the aspects could be labelled *vocalic*, that is, entailing vowels. It is a universal feature of human languages that they contain vowels in their segmental inventories. There is no known language that would use only consonants in its spoken form. The acoustic foundation of vowels rests on the fact that the raw voicing sound passing through the supralaryngeal cavities of the vocal tract induces greater vibrations at specific frequency bands. These stronger frequencies are perceptually salient and they are called *formants*. The first and second formant, marked $F1$ and $F2$ respectively, account for most differences between vowels, which makes them an elegant and parsimonious set of parameters. Vowel formants can be visualized in a two-dimensional plot and, as acoustic dimensions, they are closely related to the corresponding articulatory dimensions, that is, vowel height and the frontness or backness of the tongue. The $F1$–$F2$ plots may thus display vocalic differences between languages, language varieties, as well as individual speakers.

A modern approach to vowel formants dates back for almost two centuries. Formant specifications are used to capture vowel qualities, whether we are interested in (a) the vocalic system of a given language (e.g., Beňuš, 2012, for Slovak; Deterding, 1997, for Standard British English; Hillenbrand et al., 1995, for American English; Skarnitzl & Volín, 2012, for Czech), (b) comparing varieties of a language (e.g., Escudero et al., 2009, for Brazilian vs. European Portuguese; Ferragne & Pellegrino, 2010, for varieties of British English; Fox & Jacewicz, 2009, for varieties of American English; Hawkins & Midgley, 2005, for age-related variation), or (c) analysing vowels in foreign-accented speech (e.g., Leppik et al., 2019; Munro et al., 2013; Šimáčková & Podlipský, 2018; Šturm & Skarnitzl, 2011).

In the case of English, the specific research field of vowels must be sometimes further broken down for analytical purposes because English has a large number of vowel contrasts. Standard British English (SBE) is described to have 12 monophthongal and eight diphthongal vowel phonemes. (This arrangement should not be taken as a strict law. Due to the constant development of sound patterns in English, 13 monophthongs and seven diphthongs are suggested by more recent accounts and different numbers can be found in the description of various English accents, e.g., Ball, 1984; Hawkins & Midgley, 2005.) The reason why foreign accents matter stems from their controversial position in our lives. Although our typical explicit declarations speak about tolerance or indifference, the implicit impact of foreign-accented speech on perception (and also self-perception) of speakers is indicated by many research results (e.g., Brennan & Brennan, 1981; Bresnahan et al., 2002; Gluszek & Dovidio, 2010; Kavas & Kavas, 2008; Lev-Ari & Keysar, 2010; Lindemann, 2003; Miller & Hewgill, 1964; Rubin, 1992). It is highly irresponsible to claim that the effects of foreign-accented speech will disappear if we disregard them (see Volín, 2018).

Czech learners of English display various degrees of aptitude and they learn to speak the language with varying success. It could be said that they are familiar with the front mid /e/ and back mid /ɔː/ English sounds because these two vowels have their close equivalents in the Czech language. In contrast, the open to open-mid /æ/ and /ɒ/ are unusual sounds that they must acquire if they wish to speak without a conspicuous foreign accent. Given that the sound contrasts of /e/ vs. /æ/ and /ɔː/ vs. /ɒ/ are seemingly analogical, we set out to investigate to what extent these contrasts contributed to the perception of Czech-accented English. Inspired by a study (Torstensson et al., 2004) that asked native Swedish speakers to mimic Swedish spoken by native British English speakers, we asked a group of Czech learners of English to produce the English they considered their best self-representation and to mimic a "typical" Czech accent of English. Our premise was that Czech learners could share a view or cognitive prototype of the properties of the English-accented Czech just as the Swedes in that study did with English-accented Swedish. To this aim, we developed an experiment that assessed holistically Czech learners of English accent via external evaluators (e.g., experienced teachers of English) and internal mechanisms (i.e., the learners' judgement of best English vs. Czech accent of English). We then explored the two vowel contrasts in each condition to answer the following research question:

1. To what extent are /e/ vs. /æ/ and /ɔː/ vs. /ɒ/ likely to participate in external/internal concepts of Czech accent of English?
    a. Do Czech learners of English change their vowel contrasts when asked to enhance their Czech accent of English?
    b. Do three externally assessed subgroups differentiate between the vowel phonemes in opposition?

Another research question that is more distant yet related to our current research is that of symmetricity in phonological inventories. Descriptions of segmental systems of languages mention apparent parallelisms, for instance, between front and back, or open and close vowels. However, distributional and combinatory properties of individual vowel phonemes seldom show any mirroring. Hence, we also sought to investigate any potential analogies between the /e/ × /æ/ and /ɔː/ × /ɒ/ oppositions in our research context. Specifically, our second research question was:

2. Is there symmetricity in the students' performance, that is, do individual students treat the front and the back contrast analogically in their speech production?

The answer to this question should contribute to the solution of a larger problem concerning competing models of speech representation by human cognitive mechanisms. Symmetrical behaviour would favour deterministic models, in which the properties of the inventory guide the internal structures. Asymmetricity would favour episodic (or stochastic) models, in which the internal structures are guided by the properties of the actual usage, that is, frequencies of occurrences, repeated contextual dependencies, etc. (cf. Moore, 2007; Pisoni, 1997).

2 Method

2.1 Participants

Thirty-eight university students of various philological programmes (all female, aged 19–24) took part in the experiment. Recording their speech was a routine component of their seminar work for which they were awarded credits. They were all native speakers of Czech, and their competence in English was at the level of B2 to C1 of the CEFR. They were also fluent readers without any hearing problems or speech impairments. They were told that their recordings would be stored anonymously and later analysed, but they were not informed about the exact purpose of the experiment.

During the experiment, participants were assigned to either one of two conditions: internal ($N = 20$) and external ($N = 18$). The former was labelled "internal" (INT) because participants were asked to read the same text twice guided internally to produce less, and later more accented speech. The latter was labelled "external" (EXT) because participants' accentedness was assessed by experienced pronunciation teachers (Skarnitzl et al., 2005), who were not aware of the purpose of the experiment.

2.2 Materials and Procedure

Two news bulletin texts from recent broadcasts in English were used in the experiment. Participants were recorded as they read one of these texts aloud. The recordings took place in a sound-treated studio booth of the Institute of Phonetics at Charles University in Prague. The AKG C4500 B-BC condenser microphone was used connected with the external sound card Steinberg UR44. The signal was captured by Adobe Audition CS6 package at a sampling rate of 32 kHz and an amplitude resolution of 16 bits. Prior to recording, participants were given sufficient time to get acquainted with their texts and were invited to consult any of their possible lexical or grammatical doubts.

The INT group read aloud a shorter text (250 words) twice. They were asked to use their best English pronunciation during the first reading and imitate a typical Czech accent of English during the second reading. Their two renditions were separated by the reading of a Czech text of a comparable length to facilitate the transition to

Czech-accented English. The first reading was labelled “reduced Czech accent of English” (INT-R) and the second “enhanced Czech accent of English” (INT-E).

The EXT group was asked to read aloud a longer text (500 words; seven paragraphs) fluently using their best English pronunciation. Nine experienced teachers of English were asked to rate participants’ recordings based on a general impression of their accentedness, where 1 = *Little or No Accent*, 2 = *Ambiguous Accent* and 3 = *Strong Czech Accent*. The raters were not aware of the purpose of this experiment and were explicitly asked not to focus solely on a specific pronunciation feature but rather capture their general feeling. From the original number of 60 recordings we selected 18 with the highest agreement among raters and assigned them to one of three experimental subgroups: “inconspicuous” (EXT-1) ($n = 6$), “moderate” (EXT-2) ($n = 6$), and “strong” (EXT-3) ($n = 6$) Czech accent of English.

2.3 Analyses

After the recordings, individual instances of the four vowels under study—i.e., /e, æ, ɒ, ɔː/—were identified in the two texts phonemically with respect to SBE, i.e., regardless of the actual sound. This means that the vowel in the word *act* was categorised as /æ/ even if the speaker pronounced clear [ekt]. Similarly, the vowel in *hot* was categorised as /ɒ/ even if it sounded more like [o] or [ɑ]. This is because we were primarily interested in the realizations of the underlying (canonical) segments. However, the auditory impression was not ignored entirely (see Sects. 3.3 and 3.4). The vowel boundaries were manually positioned based on visual and auditory inspection. Only stressed vowels were measured and only those that were not aimed at a different vowel (e.g., items where *acrid* was incorrectly realized as [eɪkrɪd] or *court* as [kɜːt] were excluded).

Formant values were obtained in Praat (Boersma & Weenink, 2016) using the default robust extraction settings for female speakers (10 poles, 5500 Hz upper limit). In total, 480 vowels were analyzed in the INT group and 2145 in the EXT group. A mean value for each formant was calculated from five equidistant values extracted in the central third of each vowel. This procedure reduced the noise in the data caused by transient portions of the vocalic formants at the boundaries of vowels with consonants.

As the formant values correlate to a considerable extent with the position of the tongue in the oral cavity, we refer to vowels with higher values of $F1$ as more open and vowels with higher values of $F2$ as more front. This metonymy is common in current presentation practice, even if the two-dimensional image of the tongue does not explain the vowel acoustics in its entirety. Naturally, to make the correspondence between the sagittal cross-sections of the oral cavity and acoustic plots clear, $F2$ must be placed on the abscissa (x-axis) and $F1$ on the ordinate (y-axis), and the zero value needs to be conceptualized on the top right.

Statistical analyses of the data involved *t*-tests for repeated measures and linear mixed-effects models using the *lme4* package in R followed by calculations of Tukey post-hoc contrasts computed with the *multcomp* package in R (Bates et al., 2015; Hothorn et al., 2008; R Core Team, 2019).

3 Results

This section first presents the results for the INT and EXT groups separately. It then compares the mean formant values obtained in each group to those from a group of native speakers of English reported by Deterding (1997).

3.1 The Internal Assessment Group

The mean formant values captured in Fig. 1 illustrate the situation in the INT sample. The first obvious thing to notice is the smaller distance between the oppositions under the enhanced accent (INT-E) condition: the black targets are closer to each other both for the front pair [e, æ] and for the back pair [ɔː, ɒ]. In the case of the front vowels, the difference is caused by $F1$ while $F2$ is virtually identical for the white-black pairs. Moreover, $F1$ only matters for the open front [æ]. Generally, when Czech speakers

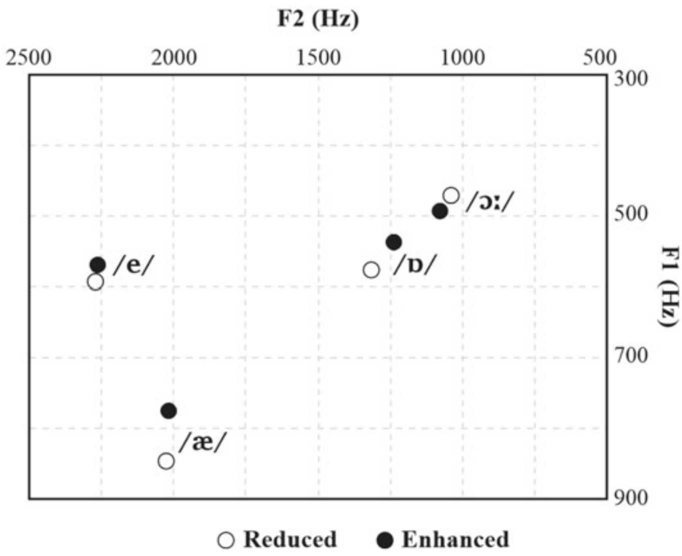

Fig. 1 The vocalic space with vowels /æ, e, ɒ, ɔː/ produced under the INT-R (reduced Czech accent) and INT-E (enhanced Czech accent) conditions

of English produce [e], the outcome is almost identical under both conditions (*t*-test for repeated measures established the differences for both formants as statistically insignificant). However, under the INT-R condition, [æ] is clearly more open: $t\ (19) = 2.59$; $p = 0.018$.

The back opposition does not display any substantial analogy, even though we might say that the mid [ɔ:] is again almost the same under both conditions (the differences for both formants are statistically insignificant), while the open-mid [ɒ] is more open under the INT-R condition. Interestingly though, *t*-tests for repeated measures found significance only for $F2$: $t\ (19) = 2.89$; $p = 0.009$. The result for $F1$ was not significant: $t\ (19) = 1.56$; $p = 0.134$.

For the purpose of generalization, however, it might be useful to check how individual speakers contribute to this sample outcome. Figure 2 displays the difference in the openness of [æ] between the INT-R and INT-E conditions. It can be observed that only 13 of the 20 speakers produced their [æ] more open when speaking with the reduced accent (INT-R condition), that is, the white column is longer than the black column in the graph. Seven others produced the opposite trend. However, if we turn to perceptual rather than statistical significance, we can evaluate the situation somewhat differently. According to past research in just noticeable differences, the formant values are perceived as different if they are at least five percent from each other (see, e.g., Pols, 1999). Of the 20 speakers, 11 produced noticeably more open [æ], five did not differentiate much, and four produced a closer [æ] under the INT-R condition. In other words, slightly more than a half of the sample indicated the knowledge (whether implicit or explicit) of the required openness of [æ] in SBE pronunciation.

Although the analogical treatment of openness in [ɒ] did not reach statistical significance (as previously shown), we examined it as we did with [æ]. This decision

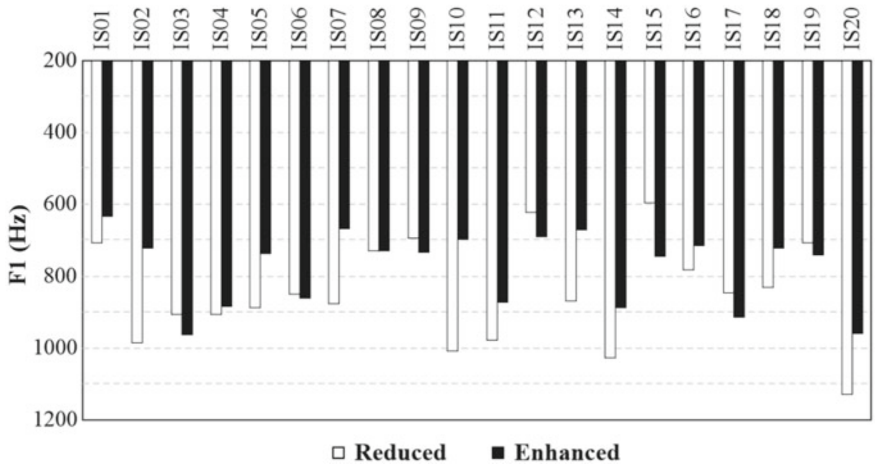

Fig. 2 The mean values of $F1$ for the vowel [æ] by individual speakers under the INT-R (reduced Czech accent) and INT-E (enhanced Czech accent) conditions

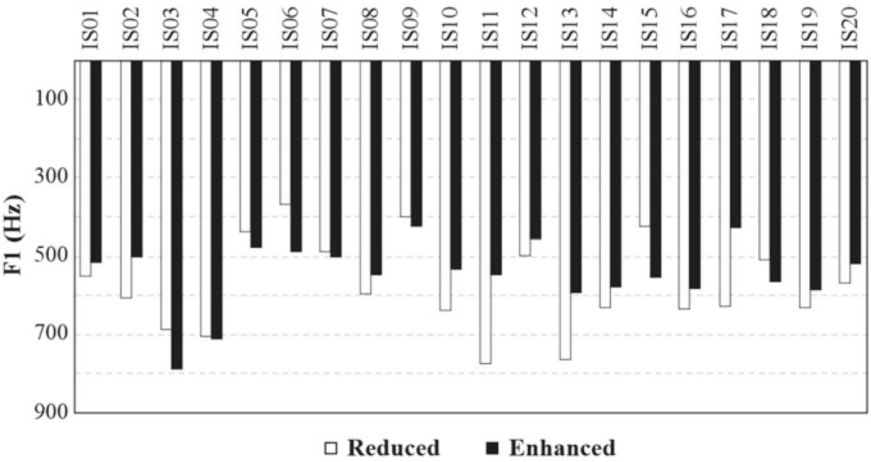

Fig. 3 The mean values of $F1$ for the vowel [ɒ] by individual speakers under the INT-R (reduced Czech accent) and INT-E (enhanced Czech accent) conditions

was motivated by the apparent phonological parallelism of /æ/ and /ɒ/ in the system of SBE vowels. Figure 3 displays the values. Similarly to the situation with [æ], 12 speakers produced more open [ɒ] under the reduced accent condition (INT-R), which is the desirable outcome. Nine of those also produced more open [æ] under INT-R (but one did not reach perceptual significance). In abstract terms, these speakers exploited the feature of openness for both the front and the back opposition in order to reduce their accents. Conversely, it is worth noticing that speaker IS15 produced the vowels consistently against the arrangement in the SBE system.

3.2 The External Assessment Group

The EXT speakers, who were externally assessed and divided into three subgroups according to their overall pronunciation competence, produced formant values that resonate to some extent with the situation in the INT group, but add some more information on Czech-accented English.

Figure 4 shows that the difference among the three EXT subgroups is smaller for [e] and [ɔ:] than for [æ] and [ɒ]. A plausible explanation for this is that the vowels [e] and [ɔ:] have their close counterparts in Czech and, therefore, Czech speakers of English are not compelled to modify them in any significant manner. The other two vowels ([æ] and [ɒ]) are more scattered in our vocalic plot, but more importantly, the phonological contrast between neighbours is emphasised in the speech of the best pronouncing subgroup (EXT-1) (white shapes) and curtailed by the worst pronouncing subgroup (EXT-3) (black shapes). The acoustic distance between [e] and [æ], and between [ɔ:] and [ɒ] is clearly the largest for the white shapes in

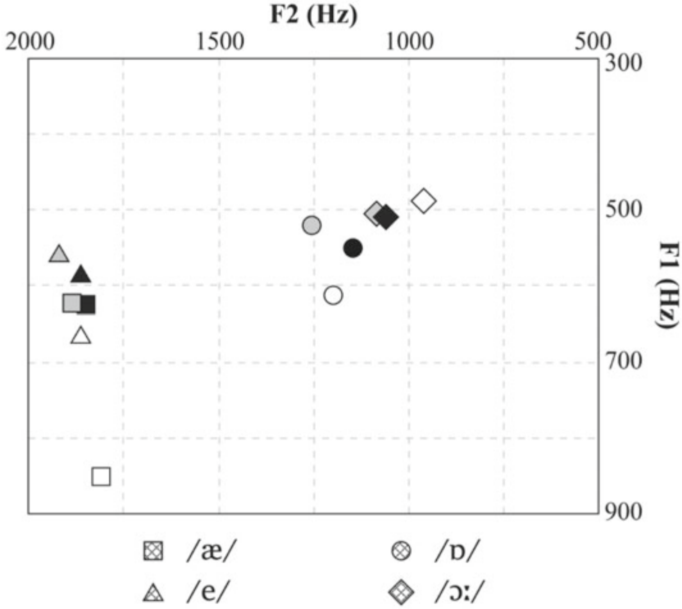

Fig. 4 The vocalic space with vowels /æ, e, ɒ, ɔː/ produced by EXT-1 (white), EXT-2 (grey) and EXT-3 (black) groups; [e]—triangles, [æ]—squares, [ɒ]—circles, [ɔː]—diamonds

the graph, that is, for the speakers who achieved the highest pronunciation scores (EXT-1). The 'black group' (EXT-3) produced quite minimal differences, and the 'grey group' (EXT-2) was only marginally better. It should perhaps be repeated at this point that the pronunciation scores were based on the overall impression, that is, they were not focused on vowels.

Since there were enough data points obtained from the larger texts read by the three EXT subgroups, linear mixed-effects models were built using the *lme4* package in R. The individual speakers were set as a random effect (SPEAKER), while PHONEME (i.e., /e, æ, ɒ, ɔː/) and LEVEL (i.e., EXT-1, EXT-2, and EXT-3) were treated as fixed effects. The p values in Table 1 are based on calculations of Tukey post-hoc contrasts (computed with the *multcomp* package in R).

Interaction between LEVEL and PHONEME was established as significant for both $F1$, $\chi^2(6) = 161.2, p < 0.001$, and $F2$, $\chi^2(6) = 31.6, p < 0.001$. Table 1 summarizes the relevant comparisons between open-mid and mid vowels at the front or back. Both contrasts were significantly different in $F1$ and $F2$ in the EXT-1 subgroup (inconspicuous accent). In the EXT-2 subgroup (moderate accent), the front vowels differed significantly only in $F1$ and back vowels in $F2$. In the EXT-3 subgroup (strong accent), $F2$ did not show any significant differences, and only the front vowels significantly differed in $F1$. However, when taking into account effect sizes, any significant effects in the EXT-2 and EXT-3 subgroups were considerably lower than in the EXT-1 subgroup. Also, openness ($F1$) tended to be associated with greater changes than frontness ($F2$).

Table 1 Pairwise differences between formants of open-mid and mid vowels in EXT subgroups. Adjusted *p*-values from Tukey contrasts in a multiple comparison of means applied to an LME model

Formant	Level	Opposition	Estimate [Hz]	*SE*	*z*	*p*
$F1$	EXT-1	/æ/ – /e/	185	13.5	13.74	< 0.001***
$F1$	EXT-1	/ɒ/ – /ɔ/	101	18.3	5.50	< 0.001***
$F1$	EXT-2	/æ/ – /e/	75	12.4	6.09	< 0.001***
$F1$	EXT-2	/ɒ/ – /ɔ/	−5	17.4	−0.29	1.00
$F1$	EXT-3	/æ/ – /e/	64	12.7	5.01	< 0.001***
$F1$	EXT-3	/ɒ/ – /ɔ/	11	17.1	0.65	0.99
$F2$	EXT-1	/æ/ – /e/	−83	21.9	−3.77	< 0.01**
$F2$	EXT-1	/ɒ/ – /ɔ/	176	29.9	5.89	< 0.001***
$F2$	EXT-2	/æ/ – /e/	−37	19.8	−1.86	0.70
$F2$	EXT-2	/ɒ/ – /ɔ/	161	27.7	5.79	< 0.001***
$F2$	EXT-3	/æ/ – /e/	−51	20.5	−2.46	0.28
$F2$	EXT-3	/ɒ/ – /ɔ/	49	27.3	1.78	0.76

EXT-1 = inconspicuous accent; EXT-2 = moderate accent; EXT-3 = strong accent
* = marginally significant (unaccounted), ** = significant, *** = highly significant

3.3 Two Sources of [æ] Values

When listening to the recordings, we noticed a potential methodological problem. Many of the EXT-2 (moderate accent) and EXT-3 (strong accent) speakers represented the phoneme /æ/ in two disparate ways. In the original Anglo-Saxon words, the common pronunciation was perceptually similar or identical to [e], but in certain internationally used words (often of Latin or Greek origin, but which also have a cognate or near-cognate in Czech), the substituting segment sounded quite like [ʌ] or [a]. Typical examples of the former set are words like *back*, *bank*, *crash*, *man*, *sack*, *tank*, while the latter set can be exemplified by *Africa*, *album*, *Castro*, *collapse*, *commander*, etc. Averaging vowels from both sets could bias the results: we can easily imagine that calculating an arithmetic mean from [e] and [a] will lead to values close to [æ]. Figure 5 shows what happens if the representations of the phoneme /æ/ are measured as three different subsets instead of just one category. The reader is invited to imagine where the average value might lie.

Two important circumstances must be stressed. First, the number of cases in individual categories were largely unbalanced: there were 218 instances of the type /æ/ → [e], but only 76 instances of /æ/ → [ʌ]. Second, the three groups of speakers contributed to the individual categories in a different proportion: the type /æ/→ [æ] was almost exclusively satiated by the speakers with inconspicuous foreign accent (EXT-1) with some small contribution from the speakers with moderate accent (EXT-2).

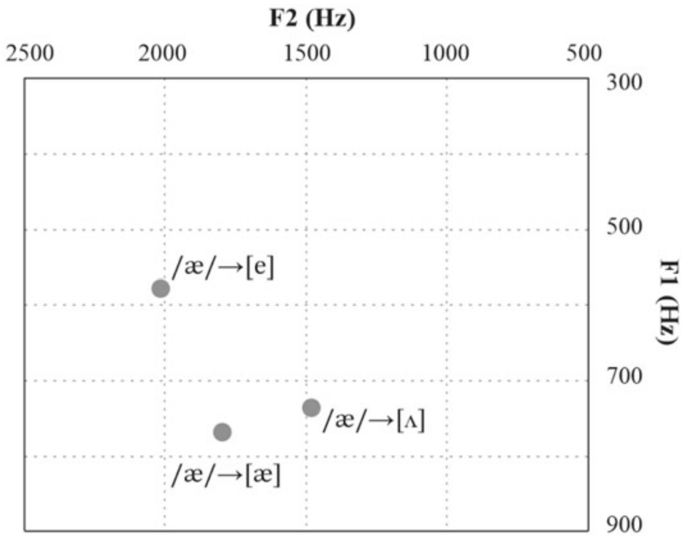

Fig. 5 The vocalic space with three perceptually distinct representations of the phoneme /æ/ found in the vowels produced by the EXT group

Table 2 Standard deviations (in Hz) from mean formant values in vowels corresponding to /e, æ/ produced by INT and EXT speakers

	F1 of /e/	*F1* of /æ/	*F2* of /e/	*F2* of /æ/
INT-speakers	92.3	121.8	126.9	148.0
EXT-speakers	82.3	124.6	126.6	231.4

INT = internally assessed; EXT = externally assessed

The inspection of standard deviation values as a correlate of variability in formant values also suggests that what we identified as [æ] because it represented the phoneme/æ/ is different from [e]. Table 2 indicates that the standard deviations are substantially higher for realizations of /æ/ than of /e/ in both groups (INT and EXT) and both vocalic formants ($F1$ and $F2$).

However, apart from two or three disparate representations of /æ/, the higher variance could also be the result of a less firm articulatory target for this foreign element. It could be argued that the target for /e/ is firmer as it is identical with a frequently occurring vowel in the speakers' L1.

Table 3 Mean values of the first two formants of vowels [e, æ, ɔ:, ɒ] under various conditions

	[e]		[æ]		[ɔ:]		[ɒ]	
	$F1$	$F2$	$F1$	$F2$	$F1$	$F2$	$F1$	$F2$
DET	719	2063	1018	1799	389	888	751	1215
INT-R	593	2269	848	2019	472	1041	575	1317
INT-E	571	2258	776	2009	491	1077	538	1234
EXT-1	664	1861	851	1805	492	967	613	1198
EXT-2	557	1920	625	1885	505	1085	519	1250
EXT-3	584	1964	623	1846	509	1069	547	1149

DET = Deterding (1997); INT-R = internally assessed, reduced accent; INT-E = internally assessed, enhanced accent; EXT-1 = externally assessed, inconspicuous accent; EXT-2 = externally assessed, moderate accent; EXT-3 = externally assessed, strong accent

3.4 Comparison with Native English

Although our primary concern in the current study is the relationship of accentedness assessment to vocalic formants in Czech speakers of English, a comparison of the obtained values with the reference values of native speakers of English found in the literature might be of interest. We took our reference values from Deterding (1997), Table 2, columns for female speakers (p. 49). Table 3 displays the summary of the mean values obtained from three groups of Czech speakers (INT-R and INT-E are actually one group under two conditions).

When we compare Deterding's (1997) reference values with our other values within each column, we can observe that in six out of eight cases the closest values to the native speaker sample are those produced by EXT-1, that is, the externally assessed speakers with inconspicuous accent. In two columns only, the closest match comes from the INT-R group: $F2$ for [e], and $F1$ for [ɔ:]. This result also corroborates the suggestion that the investigated vowels do play a certain role in impressionistic evaluations of accentedness.

4 Discussion

It is well known that vowels are somehow less phonologically charged than consonants throughout the languages of the world. They withstand greater distortions and their neutralization does not necessarily lead to a collapse in spoken interaction. Therefore, we wanted to investigate whether vocalic contrasts enter the concept of foreign accent held by external assessors or the speakers themselves, even if they are not consciously highlighted.

Our results indicate that under the internal assessment condition, our participants (Czech speakers of English) produced mid vowels [e, ɔ:] virtually the same way for both the reduced and enhanced accent, while their lower vowels [æ, ɒ] differed in that [æ] was more open, while [ɒ] was more centralized for the reduced accent. This tendency is in line with the requirements of standard pronunciation (SBE). However, individual speakers contributed to this outcome in a non-uniform manner. About one half of the speakers seemed to use the openness feature consistently and quite saliently to reduce their Czech accent, but the rest was apparently either not aware or not in control of the feature. Among the twenty speakers in the INT sample, there was also one striking exception. This speaker (IS15) used openness of [æ] and [ɒ] saliently against the reference SBE system. This result resonates with the findings of Neuhauser (2008), who concluded that her speakers were generally unable to perform consistently, and of Hao and de Jong (2016), who acknowledged that artificial production of an accent was not easy. Conclusions like these, however, should always take into account the speakers' exposure to different accents (cf. Szpyra-Kozłowska, 2018). Our sample comprised speakers who were more than familiar with the Czech accent of English—it was the most frequently heard foreign accent in their experience. Therefore, the individual variation should be attributed to personal sensitivity or acuity of the speakers.

The externally assessed (EXT) group provided data that can be reconciled with the results discussed in the previous paragraph. The subgroup with an inconspicuous Czech accent in their English (EXT-1) produced both vowel contrasts in line with descriptions of the Southern British Standard. Their open-mid vowels were, indeed, more open than their mid vowels. The subgroup with strongly accented English (EXT-3), on the other hand, made very small distinctions between the representations of /e/, /æ/ and /ɔ:/, /ɒ/, respectively.

A methodological note has to be made about averaging values for the open front vowel [æ] across the board. Our sample indicated that this could lead to serious flaws since the texts included words of Latin, Greek or other non-Anglo-Saxon origin where many Czech speakers do not even attempt at [æ]. In our analysis, we observed that the /æ/ was pronounced as [e] in English words with no equivalent in Czech, while the /æ/ was pronounced as [ʌ] or [a] (central open-mid or open vowel) in some 'international' words. This is a relatively new problem in our line of research. In the past, vowel formants were often measured in restricted sets of words that were recorded in isolation, and not in open sets of continuous meaningful texts.

A similar problem is posed by the consonantal influence on vowels. Rather than carefully pronounced sets like *heed – hid – head – had*, etc., continuous texts comprise various combinations of consonants with vowels, and their effects are not necessarily cancelled out because the phonotactic properties of vowels are not balanced in the phonological system of the English language. Specifically in our case, the back mid vowel /ɔ:/ was very often followed by the liquids /r, l/, both of which are known to pull formants in neighbouring vowels down (e.g., Volín & Studenovský, 2007). Therefore, the physical contrast we found between the pronounced [ɒ] and [ɔ:] in our sample could have been exaggerated by the liquid effects. However, this bias is part of the sound structure of English, so trying to neutralize it might make certain results less ecologically valid.

It follows that when making generalizations about foreign-accented speech, the type of speech material has to be taken quite seriously (cf. also Thompson, 1991, for a similar conclusion). Another caveat to be considered concerns the representativeness of our sample. First of all, it has to be stressed that the sample consisted of philologically oriented university students with positive attitudes to English. Therefore, it does not represent the Czech population as such. Moreover, the EXT sample was artificially balanced to comprise equal numbers of speakers in each subgroup. It can be hardly expected, however, that one third of the Czech population of learners of English is formed by speakers with an inconspicuous accent. It is quite realistic to expect that people with a strong accent prevail. Nevertheless, these considerations are not crucial for our research question. We set out to find out whether there is a link between vowel contrasts as expressed in formant measurements and criteria of general accentedness in Czech English. Our research indicates that such a link exists.

5 Implications

The results of our experiment contribute to two major areas of expertise: (a) didactics and design of classroom practice, and (b) theoretical models of speech processing. As to the first one, although the vowel qualities seem to correlate with both the internal and external evaluations of foreign accented speech, a substantial number of learners was apparently not aware of the difference. Even those who were, did not produce the contrasts strictly in parallel. Generally speaking, learners of English would most probably benefit from practicing various pairs of contrasts interlinked by a shared distinctive feature. In our case this would mean that rather than training the pronunciation of [e] vs. [æ] on its own, and sometime later focusing on the isolated pair [ɔ:] vs. [ɒ], the students should practice both contrasts together because they are both based on the feature of openness. In other words, pronunciation training should not be based on isolated phonemic contrasts but on pairs or groups of contrasts that share a salient feature.

As to the second area, the lack of strict symmetricity in our data seems to support the foundation of the episodic models of speech perception. These models maintain that rather than the abstract properties of the phonological inventory, human speech processing relies on the stochastic features of language use. If individual items of the inventory occur in different phonotactic combinations, situational contexts or types of lexicon, they will be treated differently despite the apparent analogy in the inventory arrangement. In classrooms, this would imply advantage of generous language input (namely active listening) over explanations of language structures.

6 Conclusions

We hope that our investigation, which provided formant measurements for cross linguistic comparisons, has contributed to the emerging picture of the mechanisms of foreign accents in speech. We believe that it would be beneficial to the understanding of second language acquisition to contrast the performance of Czech students of English to that of speakers of other languages. However, to draw any useful analogies, our methodology has to be replicated as closely as possible. We especially warn against using material based on pronunciation of isolated words or contextless sentences.

Our study focused on two apparently parallel phonological contrasts: /e/ vs. /æ/ and /ɔː/ vs. /ɒ/. We found out that their resonance with holistic assessments of accentedness is stronger in external conditions (in our case by experienced teachers). Internal assessments (by the speakers themselves) involved the investigated vowel contrasts in a smaller scale. It follows that if the L2 learning objectives include acquisition of sound patterns of the target language, the learners must be directed to the specific vocalic contrasts through the teaching process. Our data suggest that the problem is resolved spontaneously in only about one third to one half of the population. Other learners need external help.

Naturally, there is always the possibility to resign on pronunciation training and leave the accentedness unaltered. Unfortunately, the consequences of such a decision are still poorly understood since the research in this area is often biased by wishful thinking. What is ultimately needed are rigorously planned perception experiments informed by production studies like ours.

Acknowledgements This work was supported by the OP VVV project no. CZ.02.2.69/0.0/0.0/17_050/0008466 and by the ERDF project 'Creativity and Adaptability as Conditions of the Success of Europe in an Interrelated World' (No. CZ.02.1.01/0.0/0.0/16_019/0000734).

References

Ball, M. (1984). The centring diphthongs in Southern English: A sound change in progress. *Journal of the International Phonetic Association, 14*(1), 38–44. https://doi.org/10.1017/S0025100300002723

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beňuš, Š. (2012). Phonetic variation in Slovak yer and non-yer vowels. *Journal of Phonetics, 40*(3), 535–549. https://doi.org/10.1016/j.wocn.2012.03.001

Boersma, P., & Weenink, D. (2016). *Praat: Doing phonetics by computer* (Version 6.0.19) [Computer software]. http://www.praat.org

Brennan, E. M., & Brennan, J. S. (1981). Measurements of accent and attitude toward Mexican-American speech. *Journal of Psycholinguistic Research, 10*, 487–501. https://doi.org/10.1007/BF01076735

Bresnahan, M. J., Ohashi, R., Nebashi, R., Liu, W. Y., & Shearman, S. M. (2002). Attitudinal and affective response toward accented English. *Language and Communication, 22*(2), 171–185. https://doi.org/10.1016/S0271-5309(01)00025-8

Deterding, D. (1997). The formants of monophthong vowels in Standard Southern British English pronunciation. *Journal of the International Phonetic Association, 27*(1–2), 47–55. https://doi.org/10.1017/S0025100300005417

Escudero, P., Boersma, P., Rauber, A. S., & Bion, R. A. H. (2009). A cross-dialect acoustic description of vowels: Brazilian and European Portuguese. *Journal of the Acoustical Society of America, 126*(3), 1379–1393. https://doi.org/10.1121/1.3180321

Ferragne, E., & Pellegrino, F. (2010). Formant frequencies of vowels in 13 accents of the British Isles. *Journal of the International Phonetic Association, 40*(1), 1–34. https://doi.org/10.1017/S0025100309990247

Fox, R. A., & Jacewicz, E. (2009). Cross-dialectal variation in formant dynamics of American English vowels. *Journal of the Acoustical Society of America*, *126*(5), 2603–2618. https://doi.org/10.1121/1.3212921

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

Hao, Y. C., & de Jong, K. (2016). Imitation of second language sound in relation to L2 perception and production. *Journal of Phonetics, 54*, 151–168. https://doi.org/10.1016/j.wocn.2015.10.003

Hawkins, S., & Midgley, J. (2005). Formant frequencies of RP monophthongs in four age groups of speakers. *Journal of the International Phonetic Association, 35*(2), 184–199. https://doi.org/10.1017/S0025100305002124

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America, 97*(5), 3099–3111. https://doi.org/10.1121/1.411872

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal, 50*(3), 346–363. https://doi.org/10.1002/bimj.200810425

Kavas, A., & Kavas, A. (2008). An exploratory study of undergraduate college students' perceptions and attitudes toward foreign accented faculty. *College Student Journal, 42*, 879–890.

Leppik, K., Lippus, P., & Asu, E. L. (2019). The production of Estonian vowels in three quantity degrees by Spanish L1 speakers. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren *Proceedings of 19th International Congress of Phonetic Sciences* (pp. 1154–1158). IPA.

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

Lindemann, S. (2003). Koreans, Chinese or Indians? Attitudes and ideologies about nonnative English speakers in the United States. *Journal of Sociolinguistics, 7*(3), 348–364. https://doi.org/10.1111/1467-9481.00228

Miller, G. R., & Hewgill, M. A. (1964). The effect of variations in nonfluency on audience ratings of source credibility. *Quarterly Journal of Speech, 50*(1), 36–44. https://doi.org/10.1080/00335636409382644

Moore, R. K. (2007). Spoken language processing: Piecing together the puzzle. *Speech Communication, 49*(5), 418–435. https://doi.org/10.1016/j.specom.2007.01.011

Munro, M. J., Derwing, T. M., & Saito, K. (2013). English L2 vowel acquisition over seven years. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference* (pp. 112–119). Iowa State University.

Neuhauser, S. (2008). Voice disguise using a foreign accent: Phonetic and linguistic variation. *The International Journal of Speech, Language and the Law, 15*(2), 131–159. https://doi.org/10.1558/ijsll.v15i2.131

Pisoni, D. B. (1997). Some thoughts on 'normalization' in speech perception. In K. Johnson & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 9–32). Academic Press.

Pols, L. C. W. (1999). Flexible, robust, and efficient human speech processing versus present-day speech technology. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. C. Bailey (Eds.), *Proceedings of 14th International Congress of Phonetic Sciences* (pp. 9–16). UCLA. https://www.fon.hum.uva.nl/archive/1999/icphs/PolsICPhS1999.pdf

R Core Team. (2019). *R: A language and environment for statistical computing* (Version 3.5.3). R Foundation for Statistical Computing, Vienna. Retrieved August 26, 2020, from www.r-project.org

Rubin, D. L. (1992). Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher Education, 33*, 511–531. https://doi.org/10.1007/BF00973770

Šimáčková, Š, & Podlipský, V. J. (2018). Production accuracy of L2 vowels: Phonological parsimony and phonetic flexibility. *Research in Language, 16*(2), 169–191. https://doi.org/10.2478/rela-2018-0009

Skarnitzl, R., & Volín, J. (2012). Referenční hodnoty vokalických formantů pro mladé dospělé mluvčí standardní češtiny. *Akustické Listy, 18*(1), 7–11.

Skarnitzl, R., Volín, J., & Drenková, L. (2005). Tangibility of foreign accents in speech: the case of Czech English. In A. Grmelová, L. Dušková, & M. Farrell (Eds.), *Proceedings of the 2nd Prague Conference on Linguistics and Literary Studies* (pp. 11–20). UK PedF.

Šturm, P., & Skarnitzl, R. (2011). The open front vowel /æ/ in the production and perception of Czech students of English. In P. Cosi, R. De Mori, G. Di Fabbrizio, & R. Pieraccini (Eds.), *Proceedings of 12th Interspeech 2011* (pp. 1161–1164). ISCA. https://doi.org/10.21437/Interspeech.2011-344

Szpyra-Kozłowska, J. (2018). The rhotic in fake and authentic Polish-accented English. *Lublin Studies in Modern Languages and Literature, 42*(1), 81–102. https://doi.org/10.17951/lsmll.2018.42.1.81

Thompson, I. (1991). Foreign accents revisited: The English pronunciation of Russian immigrants. *Language Learning, 41*(2), 177–204. https://doi.org/10.1111/j.1467-1770.1991.tb00683.x

Torstensson, N., Eriksson, E. J., & Sullivan, K. P. H. (2004). Mimicked accents. Do speakers have similar cognitive prototypes? In S. Cassidy, F. Cox, R. Mannell, & S. Palethorpe (Eds.), *Proceedings of the 10th Australian International Conference on Speech Science & Technology* (pp. 271–276). ASSTA.

Volín, J. (2018). Foreign accents and responsible research. In J. Volín & R. Skarnitzl (Eds.), *The pronunciation of English by speakers of other languages* (pp. 4–18). Cambridge Scholars Publishing.

Volín, J. & Studenovský, D. (2007). Normalization of Czech vowels from continuous read texts. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 185–190). IPA a UDS. http://www.icphs2007.de/conference/Papers/1722/1722.pdf

Jan Volín is Associate Professor at the Institute of Phonetics in Prague, Czech Republic. His research focuses on sound structure of languages, general issues of speech prosody, speech acquisition, and on various pragmatic and psychological aspects of human speech behaviour. These topics alternate in his papers published in various scientific journals, chapters in monographs, and handbooks dedicated to suprasegmental phonetics.

Tanja Kocjančič Antolík's research is focused mainly on speech production in typical and atypical speech. She has been involved in different research projects across Europe addressing speech in a foreign language learning and speech disorders. She employs acoustic and articulatory analysis and is an expert in ultrasound tongue imaging.

Radek Skarnitzl is Associate Professor at the Faculty of Arts, Charles University, Prague, Czech Republic. His research focuses on second language pronunciation and the effect of various

pronunciation features on the socio-psychological evaluation of a speaker in both native and foreign languages. He is also interested in issues related to speaker identification, especially the effects of voice disguise.

Pavel Šturm is Assistant Professor at the Faculty of Arts, Charles University, Prague, Czech Republic. His research interests lie mainly in the domain of second language pronunciation (especially L2 English), focusing on segmental aspects in terms of production and on foreign accent evaluation and perception. He also examines the issues of syllabification and phonotactics in the Czech language.

On the Value of L2 Pronunciation Data for Linguistic Theory: The Story of /h/

Steven H. Weinberger

Abstract Variation remains one of the reasons why L2 speech data have not typically enjoyed the status of being linguistically motivating. This chapter deals with Mandarin Chinese learners of English who appear to do five different things when confronted with coda obstruents: aspirating the final stops, and inserting an /h/, a schwa, and voiceless high vowels [u̥] and [i̥] We show that these speech data are not only systematic, but they provide crucial evidence for the proper theoretical representation of /h/ in language. The L2 data come from Mandarin learners of English, who read monosyllabic, monomorphemic English words in a carrier phrase. Each target word ended with an obstruent and the productions were carefully transcribed. We argue that all five behaviors are instantiations of one process. This analysis illuminates the gray area surrounding the phonetic and phonological representation of /h/. Evidence from feature analyses, distribution parallels, and segment inventory implications from a range of natural languages support the claim that /h/, aspiration, and vowels are equivalent entities. Spectrographic results show that these added segments had equal phonetic durations and were significantly different from native Mandarin aspirated stops. We conclude that this apparent variation is simply a theoretically unified process of epenthesis.

Keywords V-epenthesis · Pronunciation variation · Mandarin Chinese · Coda consonants

1 Introduction

Those who study pronunciation in second language acquisition have always been plagued by the apparently fluctuating nature of the L_2 data. Certainly, a proportion of variation might be due to native language (NL) type, degree of proficiency, and a myriad of other non-linguistic factors, but there remains a type of variation that may be due to more abstract linguistic reasons. For instance, it is well-known that some

S. H. Weinberger (✉)
George Mason University, Fairfax, VA, USA
e-mail: weinberg@gmu.edu

V. G. Sardegna and A. Jarosz (eds.), *Theoretical and Practical Developments in English Speech Assessment, Research, and Training*, Second Language Learning and Teaching,
https://doi.org/10.1007/978-3-030-98218-8_12

learners of English substitute /t/ for /θ/, and other speakers substitute /s/ for /θ/. This type of variable behavior may be one of the reasons why L_2 speech data have not typically enjoyed the status of being *linguistically* motivating. That is, the data are not generally used to support linguistic theories, since most formal theories do not adequately deal with variation. However, detailed analyses have indeed shown that many types of L_2 variation are solved with some concerted linguistic research. In the case of differential substitution, it can be shown to be due to subtle and hidden influences from the NL (Lombardi, 2003; Weinberger, 1997). This chapter aims to dissolve another instance of ostensible variation in the pronunciation behavior of the L_2 of Mandarin speakers of English: epenthesis after English coda obstruents.

Native speakers of Mandarin Chinese produce English words with coda obstruents. Standard English allows the word-final singleton coda consonants shown in (1). Mandarin Chinese, on the other hand, has a much more constrained syllable structure, only allowing singleton codas (Duanmu, 2007), shown in (2).

(1) English singleton codas
[p, b, t, d, k, ɡ, m ,n, ŋ, f, v, θ, ð s, z, ʃ, ʒ, tʃ, dʒ, ɹ, l]
(2) Mandarin Chinese codas
[n, ŋ]

Since Mandarin Chinese coda possibilities are more restricted than those in English, any cursory contrastive analysis procedure would predict difficulties for Mandarin Chinese learners of English when producing English coda obstruents, and this is precisely what the L_2 literature has shown (Anderson, 1987; Broselow et al., 1998; Eckman, 1981; Hansen, 2001; Weinberger, 1994). While many of these studies reported instances of consonant deletion or coda feature changes, our concern here is focused upon vowel epenthesis (i.e., the insertion of a vowel; in this case, after coda obstruents). It is interesting to note that typically, these earlier studies documenting L_2 epenthesis merely point out that the added vowel is represented as [ə] (schwa). Yet, in this chapter we present findings from a small-scale study showing that Mandarin Chinese learners of English typically modify English coda obstruents by doing one of five things: aspirate the final stops, insert an /h/, insert a schwa, insert voiceless [i̥], or insert voiceless [u̥]. This behavior is not due to native language transfer nor is it unsystematic. We present evidence from feature analyses, distribution parallels, and segment inventory implications from a range of world languages that support the claim that /h/ and aspiration are equivalent. So far /h/ has been described as aspiration, an approximant, or most commonly, as a glottal fricative in the literature (Catford, 1977; Edwards, 1992; Hockett, 1955; Katz, 2013; Ladefoged, 2001; Laver, 1994; among others). Our L2 data not only supports the claim that /h/ and aspiration are equivalent, but also that aspirating final obstruents and epenthesizing the different types of vowels are fundamentally identical processes. Under this analysis, /h/ = aspiration = vowel. This analysis provides important research and pedagogical implications for the treatment of /h/, which has long been confusing in the phonetic literature.

The chapter starts with a description of our exploratory study and findings. Then, it presents a linguistically motivated theory of epenthetic behavior after coda obstruents based on the data obtained from Mandarin Chinese learners of English and linguistic evidence presented from feature analyses, distribution parallels, and segment inventories from other languages. It concludes with pedagogical and research implications for the treatment of /h/ after coda obstruents.

2 An Exploratory Study

This exploratory study, which focuses on the phonetic and phonological representation of /h/, attempts to investigate the phonetic range of epenthetic behavior after codas in a population of Mandarin Chinese speakers of English. It is not meant to be a statistical study, but rather a theoretically motivated phonological analysis of this variation.

2.1 Participants

Participants consisted of seven graduate students at George Mason University. They all reported that their native language was Putunghua (Mandarin Chinese).[1] They were four females and three males, and their ages ranged from 25 to 42 with a mean of 35. Their English onset age ranged from 12 to 30, with a mean onset age of 18. Their length of English residence (LOR) (USA) ranged from 1.5 years to 9 years, with a mean LOR of 4 years. Detailed demographics are shown in Table 1.

Table 1 Demographics

Subject	Birthplace	NL	Age	Onset age	Sex	Residence
S_1	Taiwan	Mandarin	28	15	male	4 years
S_2	Shanghai	Mandarin	25	12	female	3 years
S_3	Shanghai	Mandarin	38	15	male	4 years
S_4	Shanghai	Mandarin	38	15	female	1.5 years
S_5	Shanghai	Mandarin	38	30	female	1.5 years
S_6	Beijing	Mandarin	39	15	female	9 years
S_7	Beijing	Mandarin	42	25	male	4 years

[1] Even though the participants were born in various regions (Taiwan, Shanghai, and Beijing), they all reported growing up in a Mandarin Chinese environment.

2.2 Data Collection and Analysis

Participants filled out a demographic questionnaire and then completed a production task in a sound-attenuated booth at George Mason University. To complete this task, they sat at a small table with a laptop computer and a microphone situated approximately 12 inches from their mouth. The microphone was an Apogee MiC 96 k connected to the laptop recording on Audacity at 44.1 kHz 16-bit mono. The laptop screen depicted a full page with the carrier sentence on top “NOW I SAY ____,” and a word at the bottom (for instance “soup”). Only the bottom stimulus word randomly changed when the space-bar was pressed. Participants were instructed to utter the carrier phrase along with the stimulus word in a natural pace and a normal voice level. There were 115 (C)CVC stimulus words, all monosyllabic and monomorphemic. The single codas in these words consisted of 30 voiceless stops, 21 voiced stops, 34 voiceless fricatives, 19 voiced fricatives, six voiceless affricates, and five voiced affricates (see Appendix for the complete list). The recording session was preceded by an additional three training words to allow the participants to become comfortable with the procedure. The entire procedure took approximately 10 min.

The resulting data from seven participants pronouncing 115 words equaled 805 items. The recordings were phonetically transcribed by three trained transcribers. Only the coda consonant and any additional material was attended to in the transcriptions. Transcribers were required to reach consensus on all transcriptions. Initial transcriber consensus was more than 93% for all voiced coda tokens but was much lower (71%) for the voiceless coda tokens. It therefore took considerably more deliberation to reach final consensus on the voiceless coda activity.

2.3 Results

General results are shown in Table 2. There was indeed a significant amount of augmentation behavior, which for the purposes of this analysis, we will be referring to as epenthesis. The general categories of epenthesis include aspiration, voiceless vowels, and voiced vowels.

Table 2 Mean epenthetic behavior for all participants

Coda segment	Number of tokens	Type of addition	Mean %
Voiceless stop	210	Aspiration/voiceless vowel	67.6/1.4
Voiced stop	147	Voiced vowel	37.4
Voiceless fricative	238	Voiceless vowel	10.5
Voiced fricative	133	Voiced vowel	15.0
Voiceless affricate	42	Voiceless vowel	21.4
Voiced affricate	35	Voiced vowel	17.1
Total	805		28.4

Based on these transcriptions, some detailed examples of the various types of epenthesis are shown in (3):

(3) Specific epenthetic examples

a. aspiration:	soup → sou[pʰ], sou[ph][2]
b. voiced vowel:	raid → rai[də]; prove → pro[və]; fudge → fu[dʒə]
c. voiceless vowel:	loaf → loa[fu̥]; stretch → stre[tʃ ɨ̥]

We clearly found that our participants abundantly used aspiration (or /h/-epenthesis) as a favored strategy (67.6% after voiceless stops) and we also found that [ə] was the most commonly transcribed vowel for this data set (25.7% after voiced obstruents). But we cannot ignore the occurrences of the voiceless vowels—specifically the vowels shown in (3)c: the voiceless high back round vowel [u̥] and the voiceless high retroflexed vowel [ɨ̥]. These occurred after voiceless obstruents 7.5% of the time. Voiceless vowels are real entities, and while they are not found phonemically in Mandarin Chinese, nor in English, they are found in a number of world languages such as Ik, Dafla, Acoma, and in some of the Bantu languages of the Congo region (Ladefoged & Maddieson, 1990). Retroflexed vowels are indeed rare, but they can be found in the Tarascan language (Foster, 1969), and in fact even in Mandarin Chinese (Duanmu, 2007). More commonly, voiceless vowels may be surface phonetic phenomena. Indeed, English has them—particularly in the first syllables of words like *particular* and *peculiar*, and they are common in Japanese words when found between voiceless consonants (Vance, 1987).

3 Proposing a Linguistic Theory for the Treatment of /h/

We now turn to the problem at hand. In this section, we first identify the variable epenthetic behavior and then provide evidence that supports the collapsing of aspiration and /h/. Next we discuss the historical confusion with the proper phonological representation of /h/ in the literature. Finally, we provide further motivation for the equivalence of /h/, aspiration, and empty vowels.

3.1 The Variation Problem

If we simply and concretely document all of the epenthetic additions that the Mandarin Chinese participants in our exploratory study demonstrated, we would obtain a list like that found in (4):

[2] Two of the transcribers argued for a number of cases that they heard a "full /h/" in this position. We keep this representation as a valid one, that aspiration and [h] are timed similarly. For a phonological argument for this timing, see Catford (1977, pp. 115–116).

(4) 5-way variation in epenthetic behavior:

a. final obstruent aspiration[ʰ]
b. [h] insertion
c. schwa [ə] (voiced vowel) insertion
d. [u̥] (voiceless high back rounded vowel) insertion
e. [ɨ̥] (voiceless high retroflexed vowel) insertion.

Without a principled phonological theory, we are confounded by this degree of variation. If we invoke some phonology into this analysis, we should be able to collapse this list. We have already hinted that voiceless vowels are phonetically conditioned by voiceless adjacent segments in many languages such as English. A closer look at Table 2 shows that we only find the voiceless vowels after voiceless stops, fricatives, and affricates. So we can easily solve the voiced/voiceless vowel variation with a simple principle of voicing assimilation given in (5):

(5) An epenthetic vowel shares the voicing of the adjacent (coda) consonant.

Next we can attempt to account for the vowel quality differences between [ə], [u̥] and [ɨ̥]. Here we will also appeal to a theory of assimilation, in this case, place assimilation. Upon re-examining the data in (3), we see that the voiceless high back rounded vowel [u̥] is found after [f], as in *loaf* → loa[fu̥], and the high retroflexed voiceless vowel [ɨ] is found after the production of [tʃ], as in *stretch* → stre[tʃ ɨ̥]. Note that [f] is a labial consonant, and a round vowel like [u] is also considered to be labial. It appears that the feature labial is shared between the coda consonant and the epenthetic vowel. This assimilation of place features is quite common in many epenthetic phenomena across the world's languages and is well documented in loan phonologies in languages like Shona, for example (Uffmann, 2006). We can generalize this process to account for the retroflex vowel [ɨ] that follows the L2 production of the postalveolar affricate [tʃ]. Mandarin Chinese has no postalveolar consonants, but they do have a series of retroflex fricatives and affricates. We propose that our Mandarin Chinese speakers are treating the English postalveolar affricate [tʃ] as a retroflex, and thus this coda consonant is sharing its retroflex feature (place) with the epenthetic vowel [ɨ].[3] We are therefore assuming a theory of vowel underspecification (Archangeli, 1988), where the epenthetic item is simply represented as V. Now we can add the general principle of place assimilation in (6):

(6) An epenthetic vowel shares the place of the adjacent (coda) consonant.

If for some reason, (6) does not occur, this L2 grammar will simply resort to the default unstressed English vowel, [ə].[4]

[3] The data seem to show that place assimilation predominantly occurred in voiceless vowels.

[4] The default vowel in Mandarin may very well be very similar to the English schwa (see Weinberger, 1993, for details).

(7) Default V = [ə]

Now that we have principles (5), (6), and (7), we can further simplify the 5-way list in (4) to the list in (8):

(8) 3-way variation in epenthetic behavior (revision #1):

a. aspiration of final obstruent [^{h}]
b. [h] insertion
c. [V] insertion

We are still faced with a 3-way variation problem. What we find is that sometimes these L2 learners produce an aspirated stop, sometimes they are perceived as adding an [h],[5] and sometimes they add an unspecified vowel, whose ultimate surface form is determined via principles (5) and (6). Our next goal is to motivate a collapsing of (8)a and b.

3.2 /h/ Aspiration

We argue that /h/ and aspiration are merely notational variants. If we survey the ideas behind superscript diacritics represented in the International Phonetic Alphabet (IPA), we generally find that the superscript "h" does not represent a coetaneous characteristic with the segment it is superscripted to, like other superscripts such as palatalized [d^j], pharyngealized [d^ʕ], rounded [d^w], velarized [l^ɣ], etc. It rather describes a type of release that is measured from the right edge of the consonant in question. Of course, there may be some few superscript diacritics that denote such release (like nasal released consonants [d^n]), but these constitute a much smaller subset and only reinforce the apparent inconsistency with the IPA. In the following subsections we present evidence from feature analysis, inventory parsimony, and distributional parallels that justify the collapsing of /h/ and aspiration.

3.2.1 Feature Analysis of /h/ and Aspiration

According to many authors, aspirated stops are represented with the feature [+spread glottis] (Kenstowicz, 1994). Moreover, /h/ is the only segment (besides aspirated segments) that is represented with the feature [+spread glottis] (Halle & Clements, 1983; O'Grady et al., 2017). The unique sharing of this particular feature suggests that [h] and [^{h}] may indeed be equivalent items.

[5] It could be assumed that the transcriptions with the full segment [h] were due to a significantly longer augmentation of the coda, but when instrumentally measured, the forms transcribed [Ch] and [C^h] were not that different. We will nevertheless abide by the transcribers' judgements and include both transcription variants.

3.2.2 Inventory Parsimony

In 1955 Charles Hockett proclaimed that aspirated stops should be interpreted as sequences of /C/ + /h/. He based this idea by looking at languages like Korean, Sanskrit, Hindustani, and Sui. He found that all of these languages possess plain stops, aspirated stops, and /h/. According to him, "the aspiration is rather patently simply the phoneme /h/, which occurs elsewhere" (Hockett, 1955, p. 115). We can extend Hockett's idea by examining a larger language typology set. One such set is the UCLA Phonological Segment Inventory Database (UPSID), which contains information about 317 languages (Maddieson, 1984). Based upon Hockett's, 1955 study, our hypothesis should be something like the following: *If a language contains /C*^{h}*/, then it should contain /h/.* A survey of the 317 languages in the UPSID reveals that 202 of them have /h/, and 90 have aspirated stops. Of the ones that contain aspirated stops, 77 contain /h/.[6] So the hypothesis stated above is generally supported, with 77/90 (86%) of the languages surveyed having both aspirated stops and /h/.

3.2.3 Distributional Parallels

Languages that contain both /h/ and aspirated stops have been shown to exhibit a close parallel behavior in their grammatical distributions (Davis & Cho, 2003). /h/ generally occurs in the same environments as aspiration in English. Consider the aspiration data in (9):

(9) English aspiration and stress

a. C^h	b. C
a[t^h]omic	a[ɾ]om
ap[p^h]lause	ap[p]le
ca[p^h]ricious	ca[p]ricorn
re[p^h]eat	ra[p]id

As shown in (9)a, voiceless stops that are onsets in stressed syllables are necessarily aspirated in English. Those in (9)b are onsets in non-stressed syllables, and hence there is no aspiration. This is very clear in the word *atom*, where we find the common North American process of flapping. Now look at the parallel situation when we consider [h] in English onsets. These examples are given in (10).

[6] We use slashes here (/ /) to indicate that the inventories in the UPSID are typically considered to be phonemic ones. Surprisingly, Mandarin Chinese is not one of the languages in the 86%.

(10) English /h/ deletion

a. /h/	b. /h/-deletion
in[h]ibit	in[]ibition
pro[h]ibit	pro[]ibition
[h]istoric	pre[]istoric
a [h]istory of opera	an []istoric opera

Here, we see that /h/ is deleted in onset position in unstressed syllables. This behavior of /h/ in (10) is precisely the same as the behavior of aspiration in (9).

Historical evidence also supports these distributional parallels. Grassman's Law is shown to operate in Classical Greek, and states that aspiration may not appear in two successive syllables (Bidwell, 1971). So we have forms like (11), where the first instance of aspiration deletes:

(11) t^h rik^h – hair → $trik^h$os (gen. sing.)

Note that /h/ behaves similarly in (12):

(12) hed – + ap^hos → $édap^h$os *ground*, *base*

To summarize this section, we have argued that aspiration and /h/ behave the same. Therefore, we will consider them identical and arbitrarily assume that the representation is simply /h/. If this is the case, we now move from a 3-way variation in (8), to a 2-way variation in (13).

(13) 2-way variation in epenthesis (revision #2)

a. [h] insertion
b. [V] insertion.

Now if we could only collapse (13)a and b into one unified representation, we will have (perfectly) reduced the L2 epenthetic variability. This is our next task.

3.3 The Indeterminate Nature of /h/

Most introductory linguistics texts list /h/ as one of the fricative consonants, specifically a voiceless glottal fricative. Yet, there are other textbooks that appear to suggest that /h/ has the same qualities of a vowel. In this section, we survey a number of texts to paint the picture of a confusion when it comes to the representation of /h/. You will notice that the confusion with properly representing /h/ often shows up within single texts, where authors indicate multiple ways of describing this sound.

3.3.1 The Standard Version: /h/ Is a Glottal Fricative Consonant

The most common way of characterizing [h] is to group it with all the other fricative consonants. IPA charts and phonetic inventories invariably place the /h/ in the glottal place column, in the fricative row (Catford, 1977; Edwards, 1992; Giegerich, 1992; International Phonetic Association, 1999; Katz, 2013; Kent & Read, 1992; Ladefoged & Johnson, 2011; Malmberg, 1963; O'Grady et al., 2017).

3.3.2 /h/ Is a Sonorant

Some authors consider /h/ to be a sonorant glide. Chomsky and Halle assign the features [+son, -cons, -vocalic] to it (Chomsky & Halle, 1968). Edwards (1992) considers /h/ to have the features [-cons, -son]; yet he believes it is most appropriately considered to be a "hybrid glide-fricative" (p. 146). And according to Laver (1994), /h/ is a whispered approximant (see Sect. 3.4 for some distributional evidence that supports a high sonority value of /h/).

3.3.3 /h/ Has Vowel Qualities

Ladefoged and Johnson (2011) early on list /h/ in the traditional glottal fricative slot, but they later remark that "the English /h/ is the voiceless counterpart of (its) surrounding sounds" (p. 69). Although Kent and Read (1992) note that /h/ is not associated with formant transitions, it is nevertheless often totally coarticulated with a following vowel's vocal tract shape (p. 128). Malmberg (1963) also lists /h/ in the glottal fricative slot, and she alternatively calls /h/ a laryngeal spirant, referring to it as a voiceless vowel (p. 50). Edwards (1992), too, modifies his description in his text and refers to /h/ as a voiceless vowel. When introducing fricative consonants like /h/ in their book *The Sounds of the World's Languages*, Ladefoged and Maddieson (1996) state that "It is more appropriate to consider them in the chapter on vowels" (p. 137). In fact, in an earlier text, Ladefoged (2001), while listing /h/ as a fricative, says that the sound is more like a noisy (turbulent) vowel. Likewise, Reetz and Jongman (2009) discuss /h/ in their section on fricatives yet point out that /h/ may be "a voiceless variant of the accompanying vowel" (p. 29). And Rogers (2000) refers to /h/ as a phonetically voiceless vowel, even though it often functions as a consonant (p. 35). Finally, Keating (1988) refers to /h/ as less than a vowel. She regards /h/ as underspecified, and therefore transparent. She does admit that it is usually referred to as a voiceless version of a following vowel, but this is the result of feature spreading assimilation (p. 282). There is much more on the indeterminate nature of /h/, but we think Katz (2013) says it best: "The phoneme /h/ is a lost soul that needs to be given a special place of its own" (p. 101).

3.4 Distributional Evidence of the Vowelness of /h/

We are now in a position to accept some distributional and representational evidence that points to the high level of sonority for /h/. We begin by looking at a typical sonority scale and discussing some data that abide by the Syllable Contact Law (Murray & Vennemann, 1983). A typical sonority scale is given in (14), where the higher numerical value indicates higher sonority.

(14) Sonority Scale (adapted from Eckman & Iverson, 1993):

stops − fricatives − nasals − liquids − glides
0 − 1 − 2 − 3 − 4

There is also a universal principle, the Syllable Contact Law, proposed by Murray and Vennemann (1983), given in (15).

(15) Syllable Contact Law:
Given the syllabic structure A] σ σ [B, where A and B have sonority values *a* and *b*, grammars prefer that the integer resulting from *a* minus *b* be as great as possible.

To demonstrate this preference, consider the two forms in (16):

(16) a. *arna*
b. *anra*

One of them is more preferable than the other. According to Clements (1992), most humans will choose (16)a. This is because within the coda-onset sequence in (16)a, *r–n*, the sonority values are 3 and 2, respectively. So according to the first part of the Syllable Contact Law in (15), the arithmetic value is $3 - 2 = 1$. For *anra*, the example in (16)b, the sonority values of $n - r$ are 2 and 3, respectively. In this case, the arithmetic value is $2 - 3 = -1$. Since $1 > -1$, humans will prefer *arna* to *anra*.

The principle in (15) also explains why the English forms in (17) are more favored than the forms in (18) (Dineen & Miller, 1998; Pinker, 1994).

(17) razzle-dazzle (boundary distance = 3)
super-duper (boundary distance = 3).
(18) *dazzle-razzle (boundary distance = 0)
*duper-super (boundary distance = 2).

This is because the forms in (17) have a liquid-stop sequence at the syllable boundary, yielding a value of $3 - 0 = 3$ for both forms. Now contrast this with the forms in (18). These forms are relatively less desirable because their coda-onset boundary distances are smaller than their respective partner forms in (17). Now what about forms that contain /h/?

(19) harem-scarem (boundary distance = 1)
hob-nob (boundary distance = −2).
(20) *scarem-harem (boundary distance = 1, if /h/ is a fricative)
*nob-hob (boundary distance = −1, if /h/ is a fricative).

Let us start with the fact that English listeners prefer the forms in (19) rather than in (20). But if /h/ is considered a fricative consonant as portrayed in Sect. 3.3.1, then at best, there is no way to choose *harem-scarem* over **scarem-harem*, since they have equivalent values, and at worst, *nob-hob should be preferred to hob-nob. But what if /h/ was considered to be more sonorant? In order to insure the preferences in (19), we must consider /h/ to (at least) be a glide, as shown in (21):

(21) *scarem-harem (boundary distance = −2, if /h/ is a glide)
*nob-hob (boundary distance = −4, if /h/ is a glide).

This supports the claims given in Sect. 3.3.2; yet it does not conflict with the claims made in Sect. 3.3.3. Clearly, /h/ is not treated as a fricative consonant by English speakers. We suggest that it has relatively high sonority.

3.4.1 Notation and Distribution of Aspiration and V

We have argued that aspiration and /h/ are considered the same thing. In this subsection, we present notational and distributional evidence that aspiration is equivalent to a vowel. Laver (1994, pp. 348–349) considers that aspiration is actually a voiceless instance of the following vowel. According to him, the English representations in (22)a could also be accurately rendered as those in (22)b:

(22) Notational variants

a. pat [pʰæt] b. *pat* [p$^{\text{æ̥}}$æt]
pet [pʰɛt] *pet* [p$^{\text{ɛ̥}}$ɛt]

We can also infer a parsimonious assimilation in voicing and a concomitant alternation between aspiration and V in the following phonetic representation of French words. Tranel (1987) offers these forms in (23) as he reports on coda release by French speakers. Note the superscript schwa in the last example.

(23) French coda release

[kapʰ] *cape*
[sakʰ] *bag*
[bagᵊ] *ring*

Finally, the behavior of some consonant clusters in Cambodian suggests that aspiration and superscript schwa (a vowel) alternate predictably. This is shown in the examples in (24) taken from Huffman (1972).

(24) Cambodian onset clusters

a. [ksac] *sand*	b. [kʰcəj] *to borrow*	c. [kᵊɓaal] *head*
[psaa] *market*	[pʰkaa] *flower*	[pᵊɗəj] *husband*

In (24)a, we can assume that onset clusters are allowed in Cambodian, as long as one member of the cluster is a fricative. However, when the cluster contains two stops, a sound is inserted between the onset consonants. If the cluster has no voiced sounds, as in (24)b, the result is aspiration. If the cluster contains a voiced consonant, as in (24)c, the inserted sound is a superscript schwa.

To summarize to this point, we have argued that /h/ and aspiration are the same. We have also motivated the idea that /h/ has higher sonority than traditionally thought, and that there are numerous authors that suggest its vowel status. This last section suggests that distributionally and notationally, aspiration can be considered to be a vowel. Therefore, by the process of transitivity, /h/ = V. We are now prepared to revise the 2-way variation in L_2 epenthesis in (13) to a single process shown in (25).

(25) L2 epenthesis (final revision)
[V] insertion.

4 Discussion of Findings from Exploratory Study

The aim of this chapter was to theoretically account for the type of epenthetic variation that occurred when Mandarin Chinese learners of English pronounced English words with single obstruent codas. We began with a 5-way variation, and eventually reduced that to a single unified process given in (25).

We can now return to the data from our Mandarin Chinese participants. It may be useful to look a bit closer at some of their aspiration and vowel insertion behavior. If our participants were truly treating aspiration and vowel insertion as equivalent processes, then we should be able to demonstrate this uniformity. To this end, we measured the acoustic duration of the material added to the English codas in a set of samples. We used Praat (Boersma & Weenink, 2018) to do this analysis. Figure 1 shows the acoustic image for a Mandarin Chinese speaker's English production of the word *scribe*, transcribed by our team as scri[bə]. Figure 2 shows the acoustic image for a Mandarin Chinese L_2 production of the word *ripe*, transcribed by our team as ri[pʰ].

The highlighted area on the oscillogram corresponds to the added vowel in the word *scribe*. The measured duration is 75.47 ms. Now compare this duration to the measured length of the added aspiration in Fig. 2.

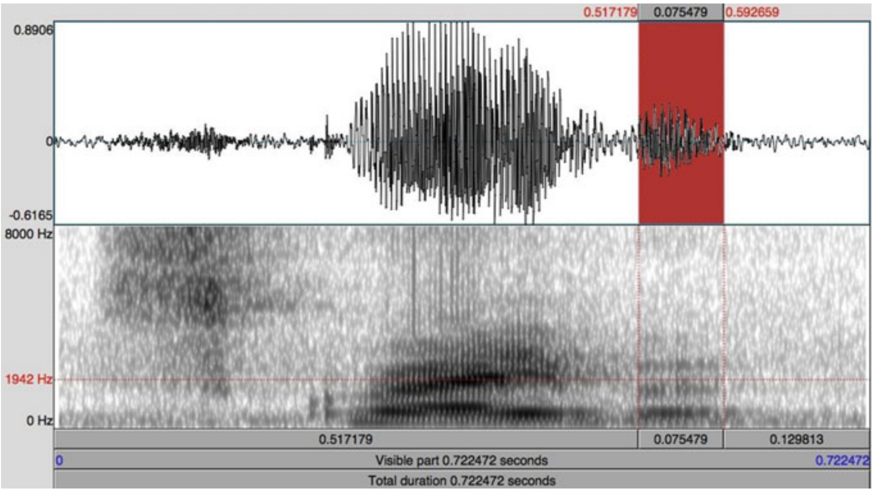

Fig. 1 The word *scribe*

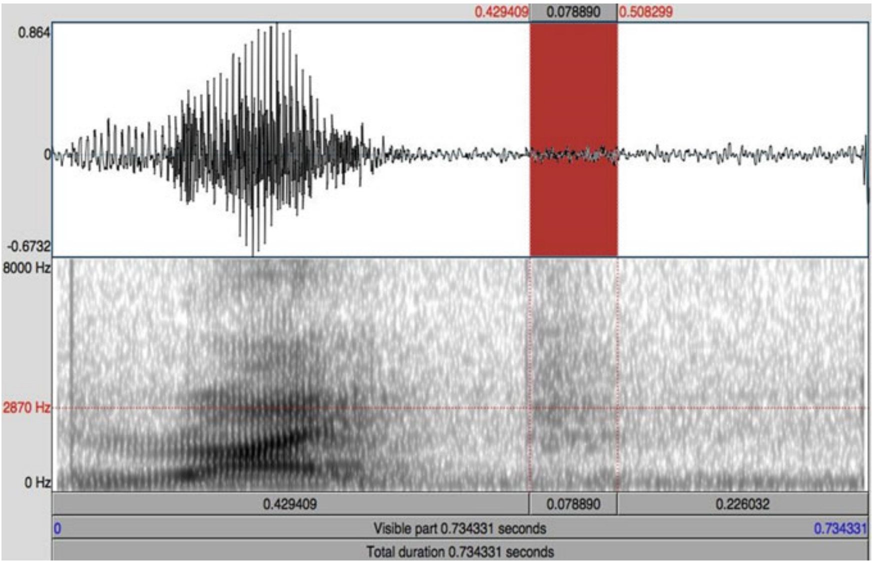

Fig. 2 The word *ripe*

The highlighted area on the oscillogram in Fig. 2 corresponds to the added aspiration in the production of the word *ripe*. The measured duration here is 78.89 ms. It seems that the durations of the added vowel and the added aspiration are very similar. We take this as added support for the unified analysis of L2 epenthesis.

Table 3 Epenthesis durations in milliseconds

Voiced codas	Epethentic vowel (ms)	Voiceless codas	Epethentic aspiration (ms)
scribe	75.47	*ripe*	78.89
abe	67.40	*pup*	71.54
tub	70.73	*soup*	56.93
M	**71.20**	***M***	**69.12**
kid	84.79	*out*	91.53
sad	82.86	*sight*	107.04
M	**83.82**	***M***	**99.28**
big	67.36	*kick*	87.31
log	99.76	*soak*	92.57
M	**83.56**	**M**	**89.94**
Total *M*	**79.53**	**Total *M***	**86.11**

There is of course some duration variation, which is due to the place of articulation of the preceding consonant. Table 3 shows the individual and mean durations for epenthetic vowels (following voiced codas) and for epenthetic aspiration (following voiceless codas). Again, the values seem to suggest a high degree of similarity between V and aspiration.

Whenever L_2 data is presented and discussed, one should remain cognizant that simple native language transfer may be playing a role. Thus, we should point out here that according to the segment inventory of Mandarin Chinese, phonemically it has no voiced obstruents but rather divides its voiceless stops into an aspirated series and an unaspirated series. Could it be that these Mandarin Chinese L_2 speakers are simply using their NL values for aspiration when they encounter English voiceless codas? We looked at published values for Mandarin aspirated stops when found in onset positions. The mean duration (VOT) for Mandarin Chinese aspirated stops is 123.14 ms. (Chao, 1992). This value seems much longer than the values reported in Table 3. We therefore dismiss any significant influence of NL transfer here. Indeed, our data actually show that these Mandarin Chinese L_2 speakers are somehow treating English voiced obstruents differently from voiceless obstruents, which is due to something other than NL transfer. Recall that they have no NL experience with obstruent codas, nor with voiced obstruents in general. When they encounter a coda obstruent in their L_2, they perform the simple and uniform epenthesis operation shown in (25) and the general assimilation principles then operate.

5 Pedagogical and Research Implications

While this research prudently makes no claims about direct applications for teachers of English, there are certainly general implications for teachers who deal with non-native speakers. What we have argued here is that Mandarin Chinese learners of English, from a surface analysis, appear to be behaving in a variable manner. And to some observers, this may convey an aspect of random pronunciation. But we have suggested that these learners are instead performing in a rather systematic fashion. They show evidence of a more uniform internal grammar of epenthesis, and the various outcomes are quite predictable. We believe that this is encouraging news for teachers, who must constantly make sense of learner variation.

From a research perspective, we have found that an analysis of epenthesis through careful phonetic transcription yields theoretically interesting results. Nevertheless, more work needs to be done with the proper representation of the voiceless vowels, as these remain the only items that appear to assimilate with respect to place of articulation in our data.

6 Conclusion

The analysis presented in this chapter has utilized accepted phonological theory to dissolve the impression that L_2 pronunciation data is unconstrained and overly variable. In dealing with L_2 coda production we have collapsed five different epenthetic behaviors—that is, final obstruent aspiration, [h] insertion, schwa [ə] (voiced vowel) insertion, [u̥] (voiceless high back rounded vowel) insertion, and [ɨ̥] (voiceless high retroflexed vowel) insertion—into one simple, unified process: [V]-insertion. In so doing, we have contributed to the proper representation of the segment /h/. Indeed, we hope we have demonstrated that L_2 data can take its place along with other linguistic data to contribute to this proper analysis of epenthetic processes.

Appendix: Word List

moose – stuff – scribe – prize – soup – teach – out – proof – freeze – ripe – peace – coach – cut – tub – said – buzz – soak – nose – sight – sad – noise – kick – dish – pitch – crush – such – pup – choice – drive – abe – lab – rope – house – kid – save – tube – weave – lobe – move – clothe – lake – kiss – mate – cloth – love – load – raise – stretch – tap – sheep – log – suck – live – rice – chef – face – catch – leash – let – big – lip – hug – loaf – path – ash – wreath – smith – trough – breathe – pit – soothe – raid – rod – oath – proud – pat – lock – neat – neck – suit – note – knife – peg – shape – lodge – pack – strafe – crib – tooth – quiz – rag – beef – fudge – ride – leak – news – stiff – squash – south – as – cook – wove – seed – food – fuss – flesh – mess – mass – rage – loss – page – breath – mop – bridge – pep.

References

Anderson, J. (1987). The markedness differential hypothesis and syllable structure difficulty. In G. Ioup & S. Weinberger (Eds.), *Interlanguage phonology* (pp. 279–291). Newbury House Publishers.

Archangeli, D. (1988). Aspects of underspecification theory. *Phonology, 5*(2), 183–207. https://doi.org/10.1017/S0952675700002268

Bidwell, C. (1971). Aspiration and /h/ in Greek and Proto-Indo-European. *Studies in Linguistics, 22*, 21–24.

Boersma, P., & Weenink, D. (2018). *Praat: Doing phonetics by computer* (Version 6.0.40) [Computer software]. http://www.praat.org

Broselow, E., Chen, S., & Wang, C. (1998). The emergence of the unmarked in second language phonology. *Studies in Second Language Acquisition, 20*(2), 261–280. https://doi.org/10.1017/S0272263198002071

Catford, J. (1977). *Fundamental problems in phonetics*. Indiana University Press.

Chao, H.-J. (1992). *Aspiration in Chinese* (Publication No. 9236418) [Doctoral dissertation, University of Illinois at Urbana-Champaign]. ProQuest Dissertations Publishing.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. Harper and Row.

Clements, G. (1992). The sonority cycle and syllable organization. In W. Dressler, H. Luschutzky, O. Pfeiffer, & J. Rennison (Eds.), *Phonologica 1988* (pp. 63–76). Cambridge University Press.

Davis, S., & Cho, M.-H. (2003). The distribution of aspirated stops and /h/ in American English and Korean: An alignment approach with typological implications. *Linguistics, 41*(4), 607–652. https://doi.org/10.1515/ling.2003.020

Dineen, N., & Miller, D. (1998). The derivation of a sonority hierarchy from the syllable contact law and the productivity of the SCL in American English. *GMU Working Papers in Linguistics, 5*, 19–47.

Duanmu, S. (2007). *The phonology of standard Chinese*. Oxford University Press.

Eckman, F. (1981). On the naturalness of interlanguage phonological rules. *Language Learning, 31*(1), 195–216. https://doi.org/10.1111/j.1467-1770.1981.tb01379.x

Eckman, F., & Iverson, G. (1993). Sonority and markedness among onset clusters in the interlanguage of ESL learners. *Second Language Research, 9*(3), 234–252. https://doi.org/10.1177/026765839300900302

Edwards, H. (1992). *Applied phonetics: The sounds of American English*. Singular Pub. Group.

Foster, M. (1969). *The Tarascan language*. University of California Press.

Giegerich, H. (1992). *English phonology*. Cambridge University Press.

Halle, M., & Clements, G. (1983). *Problem book in phonology*. MIT Press.
Hansen, J. (2001). Linguistic constraints on the acquisition of English syllable codas by native speakers of Mandarin Chinese. *Applied Linguistics, 22*(3), 338–365. https://doi.org/10.1093/applin/22.3.338
Hockett, C. (1955). *A manual of phonology*. Waverley Press.
Huffman, F. (1972). The boundary between monosyllable and disyllable in Cambodian. *Lingua, 29*, 45–66. https://doi.org/10.1016/0024-3841(72)90006-X
International Phonetic Association. (1999). *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge University Press.
Katz, W. (2013). *Phonetics for dummies*. Wiley.
Keating, P. (1988). Underspecification in phonetics. *Phonology, 5*(2), 275–292. https://doi.org/10.1017/S095267570000230X
Kenstowicz, M. (1994). *Phonology in generative grammar*. Blackwell.
Kent, R., & Read, C. (1992). *The acoustic analysis of speech*. Singular.
Ladefoged, P. (2001). *Vowels and consonants*. Blackwell.
Ladefoged, P., & Johnson, K. (2011). *A course in phonetics* (6th ed.). Wadsworth.
Ladefoged, P., & Maddieson, I. (1990). Vowels in the world's languages. *Journal of Phonetics, 18*(2), 93–122. https://doi.org/10.1016/S0095-4470(19)30396-1
Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Blackwell.
Laver, J. (1994). *Principles of phonetics*. Cambridge University Press.
Lombardi, L. (2003). Second language data and constraints on manner: Explaining substitutions for the English interdentals. *Second Language Research, 19*(3), 225–250. https://doi.org/10.1177/026765830301900304
Maddieson, I. (1984). *Patterns of sounds*. Cambridge University Press.
Malmberg, B. (1963). *Phonetics*. Dover.
Murray, R., & Vennemann, T. (1983). Sound change and syllable structure in Germanic phonology. *Language, 59*(3), 514–528. https://doi.org/10.2307/413901
O'Grady, W., Archibald, J., Aronoff, M., & Rees-Miller, J. (2017). *Contemporary linguistics* (7th ed.). Bedford.
Pinker, S. (1994). *The language instinct*. Morrow.
Reetz, H., & Jongman, A. (2009). *Phonetics*. Wiley-Blackwell.
Rogers, H. (2000). *The sounds of language*. Longman.
Tranel, B. (1987). *The sounds of French*. Cambridge University Press.
Uffmann, C. (2006). Epenthetic vowel quality in loanwords: Empirical and formal issues. *Lingua, 116*(7), 1079–1111. https://doi.org/10.1016/j.lingua.2005.06.009
Vance, T. (1987). *An introduction to Japanese phonology*. SUNY Press.
Weinberger, S. (1993). The ins and outs of empty vowels: Deletion and insertion in Mandarin. In M. Bernstein (Ed.), ESCOL 92: *Proceedings of the 9th Eastern States Conference on linguistics* (pp. 262–271). Cornell University.
Weinberger, S. (1994). Functional and phonetic constraints on second language phonology. In M. Yavaş (Ed.), *First and second language phonology* (pp. 283–302). Singular.
Weinberger, S. (1997). Minimal segments in second language phonology. In A. James & J. Leather (Eds.), *Second language speech: Structure and process* (pp. 263–311). De Gruyter Mouton. https://doi.org/10.1515/9783110882933.263

Steven H. Weinberger is Associate Professor of Linguistics in the Department of English at George Mason University, USA. He earned his Ph.D. in Linguistics in 1988 from the University of Washington in Seattle and has taught at George Mason since 1989. He teaches graduate and undergraduate level courses in phonetics, phonology, second language acquisition, and exceptional phonologies. His research deals with language sound systems, exceptional phonologies, adult second language learning, and foreign accents. He is co-editor of *Interlanguage Phonology* (1987), and is the principal investigator and administrator of the *Speech Accent Archive*.

Consistency in the Rhoticity of Czech Speakers of English

Ondřej Fischer and Pavel Šturm

Abstract Native English varieties differ greatly in the realization and distribution of the rhotic phoneme /r/. In rhotic accents, etymological /r/ is realized in all positions (rail, far, barn), whereas in non-rhotic accents its occurrence is restricted to prevocalic positions (rail) within the word. Rhoticity is one of the most salient features of English accents, and it is therefore crucial for learners of English to resolve the issue of (non)rhoticity in their L2 English production. Our main research question concerns the consistency of the speakers: If they aim for a (non-)rhotic accent, to what degree is their spoken production (non)-rhotic? We surveyed the pronunciation of 24 Czech learners of English, differing in the level of proficiency and in the preferred pronunciation model (rhotic vs. non-rhotic). The results of an auditory analysis confirmed a persistent presence of /r/ in all positions for less proficient speakers, whereas advanced learners demonstrated a lower degree of consistency and an inclination to the preferred accent model. However, although group results may suggest inconsistent treatment of rhoticity, the production of /r/ was in fact very consistent at the individual level, often attributable to background information from the participants. Pedagogical implications are discussed.

Keywords Rhoticity · Czech English · Foreign accent · Accent variation

1 Introduction

Accents are a fascinating, dynamic phenomenon, undergoing changes in time, engaging in social meaning, and growing—the number of English accents inevitably increases alongside the ever-larger community of English speakers around the globe. Accents differ in a multitude of particular aspects, but one feature can immediately be distinguished in all of them: rhoticity. It is based on a simple opposition between presence and absence of a phoneme in specific contexts, and it attracts all the more attention since the standard accents of the two most influential English-speaking

O. Fischer · P. Šturm (✉)
Institute of Phonetics, Charles University, Prague, Czech Republic
e-mail: Pavel.Sturm@ff.cuni.cz

V. G. Sardegna and A. Jarosz (eds.), *Theoretical and Practical Developments in English Speech Assessment, Research, and Training*, Second Language Learning and Teaching,
https://doi.org/10.1007/978-3-030-98218-8_13

cultures in today's world, British and American, diverge in this very aspect. Rhoticity as an accent feature is so elementary and crucial that it deserves to be thoroughly investigated in relation to learning English as a foreign language.

The phonotactic distribution of /r/ differs greatly across native and non-native varieties of English, resulting in a division of accents into *rhotic* and *non-rhotic*. The degree of rhoticity is determined not only geographically, but also socially and individually (Becker, 2014; Labov, 1966; Reed & Levis, 2015; Schneider et al., 2004; Wells, 1982). For instance, Singapore English has been traditionally listed among the non-rhotic varieties of English, and the absence of non-prevocalic /r/ is still considered "usual" in Singapore (Deterding, 2007, p. 21). However, research conducted in this field during the past thirty years indicates that there has been a certain tendency towards rhoticity. Hiang and Gupta (1992) recorded 21 Singaporean speakers from diverse social backgrounds in three speech styles (interview, read passage, read word-list). Higher usage of post-vocalic /r/ correlated with more formal styles, lower age, female sex and identification with a peer group. While the study admits the sample was not representative of the entire English-speaking Singaporean population, it reports a possible beginning of a sound change, brought about to a great extent by the growing influence of the American culture. A similar observation is made by Deterding, who claims rhoticity to be "increasingly common" in Singapore, while mentioning prestigious and cultural reasons, too (Deterding, 2007, p. 21).

In terms of language learning, rhoticity must be resolved at some point, especially if the learner strives for a particular model accent. In many European countries, Standard Southern British English (SSBE)—a non-rhotic variety—is often preferred by secondary-school and university learners (Carrie, 2017; Dalton-Puffer et al., 1997; Jakšič & Šturm, 2017; Ladegaard & Sachdev, 2006; Waniek-Klimczak & Klimczak, 2008), whereas in many Asian countries, General American (GA)—a rhotic variety—may be preferred (Chiba et al., 1995; Hansen Edwards, 2016; Young & Walsh, 2010). However, the picture is more complicated, and a number of factors influence the choice of a pronunciation model, from communicative criteria and intelligibility to pedagogical criteria and teachers' views (Mompeán, 2008).

An important consideration regarding rhoticity is the learner's native language. First, rhotic sounds display a wide range of phonetic variation both within and across languages (Ladefoged & Maddieson, 1996). One of the goals of a learner of English will thus be to acquire the target pronunciation of the rhotic as an approximant [ɹ] (and not a trill, alveolar flap or uvular fricative). Second, L2 learners of English tend to be rhotic if their L1 language allows syllable-final /r/ (Brown, 2015). In other words, the choice between GA and SSBE might depend on the phonotactic structure of the L1 language. For instance, Czech is a rhotic language, as /r/ occurs in all kinds of environment (including syllable nuclei), and GA would thus be a good option. There would also be no discrepancy between an orthographic < r > and what is pronounced. In contrast, German or Austrian learners can more straightforwardly accommodate the non-rhoticity of SSBE because German is similar in that respect. However, if a Czech learner opts for SSBE, he or she would need to master both the phonetic realization of the rhotic consonant *and* its unexpected distribution. We therefore predict that, for Czech learners, non-rhoticity will be more difficult to acquire than rhoticity.

Czech English has been described from various perspectives (e.g., Šimáčková, 2003; Skarnitzl & Rumlová, 2019; Skarnitzl & Šturm, 2016, 2017; Šturm & Skarnitzl, 2011; Volín & Johaníková, 2018; Volín et al., 2015) but rhoticity seems to be a feature that goes largely unexplored. Several university theses at the BA and MA level have been written that explicitly focus on rhoticity. Kostelná (2005) examined 20 graduates or students of English and American studies reading a text of 40 semantically isolated sentences. The overall degree of rhoticity was 67%, but varied across individuals. Interestingly, although most subjects were predominantly rhotic or non-rhotic, a few individuals confirmed that an extended stay in the target country (well over a year) can lead to a fully consistent rhotic or non-rhotic pronunciation. We should also mention the work of Pankovicsová (2014), who investigated rhoticity in Slovak English, as the Czech and Slovak languages are closely related. Slovak secondary-school students and their teachers of English read a short passage in the L2. The pronunciation of the students was clearly rhotic (99%), but this cannot be said of the teachers (63% on average). Nevertheless, the teachers differed in consistency: some were associated with varying degrees of rhoticity, others were almost always rhotic.

Several factors affect the degree of rhoticity in English. Rhotic productions have been shown to be higher in stressed syllables than in unstressed syllables and in some phonological contexts than others, defined by preceding vowel and word position (Becker, 2014; Kuecker et al., 2015; Piercy, 2012). Furthermore, there is substantial word-specific variation, variation based on the speaking style and situation, and socially conditioned variation (Asprey, 2007; Becker, 2014; Marsden, 2017; Stuart-Smith et al., 2014). Some of these L1-based facts also apply to L2 speech. Sundkvist and Gao (2016) investigated rhoticity in Chinese speakers from the Yunnan province; they reported different degrees of rhoticity across three tasks, with stronger rhoticity in more formal styles. Kang (2013) found that external factors like residence in an English-speaking country had stronger influence on rhoticity than internal, linguistic factors. This has significant implications. For non-native speech, the predominant accent in the learner's environment can affect rhoticity substantially, as in speakers consuming American culture versus those being exposed more to British input. This background information must therefore be included in a rhoticity analysis. Interestingly, Stuart-Smith et al. (2014) even mention an influence of London TV on the rhoticity of native Scottish speakers.

As hinted above, rhotic tokens may be advantageous for the English learner in that more information (an additional sound) is included in the word, facilitating bottom-up processing. Greater correspondence to written words may also help. Experimental results supporting this argument are available for Czech learners. Kobák (2017) investigated the degree of intelligibility (writing task) and comprehensibility (rating task) of rhotic and non-rhotic tokens in a perceptual experiment presented to Czech secondary-school students. Across several levels of proficiency, non-rhotic tokens were consistently associated with worse intelligibility and lower comprehensibility than rhotic tokens.

2 The Study

Our main research question concerns the speakers' consistency. We believe that rhoticity is a natural feature of Czech English. It will be strongest and most consistent in beginning learners, who rely to a large degree on the L1 patterning, and in intermediate or advanced learners socially and linguistically attracted to a rhotic variety (e.g., American English). Such learners may be informed about the choice, that is, possess knowledge about the variation in native and non-native English accents and the sociolinguistic implications of their use, or may simply have a personal preference for a variety without being aware of the wider picture. In contrast, learners with a preference (conscious in the above sense or not) for a non-rhotic variety are predicted to be less consistent in their use of rhoticity because they have to deal with the conflicting pressures of the L1, favouring rhoticity, and their preferred model, which is non-rhotic. Finally, there might also be differences between beginning and advanced learners in the phonetic realization of /r/: the less proficient speakers are expected to show a higher number of non-standard realizations (flapped [ɾ] or trilled [r]).

3 Methods

3.1 Participants

The experimental group comprised 24 Czech learners of English (20 female and 4 male; *Mdn* age = 20.5 years old). Eight speakers were learners with generally lower pronunciation proficiency (around B2), who did not study English as a university programme. Their age ranged between 20 and 59 years old (*Mdn* = 30 years old). In contrast, the remaining 16 speakers were university students of English (i.e., the English language, literature and culture) with advanced proficiency (C1–C2), including their pronunciation. Commencing their studies, their age ranged between 19 and 25 years old (*Mdn* = 20 years old). Another difference between the groups consists in the age of onset of learning, which was earlier in the more advanced group than in the intermediate group (*Mdn* AO = 8 years old versus 18.5 years old, respectively). Henceforth, the speakers will be referred to as high-proficiency (HP) and low-proficiency (LP) group. Moreover, two female native speakers were recorded as controls (one American, one British). The American speaker used a rhotic variety of English, whereas the British speaker used a non-rhotic variety.

3.2 Materials and Procedure

The speakers were recorded in a sound-treated studio reading a series of news bulletins (462 words corresponding to 3–5 min of reading time). They had sufficient time (a few minutes) for silent or semi-silent preparation in order to prevent dysfluencies. A condenser microphone was used, recording the signal as uncompressed wav files (at 32-kHz sampling frequency and 16-bit depth). The participants were unaware that they were being assessed on whether they were predominantly rhotic or not. When the recording session ended, the participants filled a short questionnaire related to accents and their language background (see Appendix). Three questions investigated language learning (length of learning English, their estimated level of proficiency in English, and other languages with a C1 or C2 proficiency level). Crucially, two other questions concerned (a) which accent of English was the most pleasant to them and (b) which, if any, of these accents they attempted to emulate, as a model accent. Another section concerned the participant's experience with English: contact with English native speakers as teachers and as acquaintances, extended stay (over a month) in an English-speaking country, and watching preferences, if any, of TV shows and series in English (a list of the most popular ones was provided, but could be expanded with their own titles).

The answers were coded in the following way. The model accent and the most pleasant accent was classified as "rhotic" versus "non-rhotic" (or coded as "none" if there was no preference). Native teachers, native speakers, and extended stay was coded as "none," "rhotic country" (i.e., USA, Canada, Scotland), "non-rhotic country" (England, Wales, India), or "balanced" (both rhotic and non-rhotic country is relevant). Similarly, their English TV input was coded as "balanced" (equal number of British and American shows or at most a difference of one), "rhotic," or "non-rhotic."

Read material was employed because conversational speech would not yield identical contexts (target words) for all speakers. In total, there were 74 target words per speaker, featuring potential /r/ in syllabic codas (31 cases) or vowel nuclei (43 cases). The majority of the former contexts involved stressed positions (e.g., "charge" pronounced as [ˈt͡ʃɑːdʒ] × [ˈt͡ʃɑːrdʒ]), whereas most of the latter contexts were associated with unstressed positions (e.g., "chapter" pronounced as [ˈt͡ʃæptə] × [ˈt͡ʃæptɚ]). However, this is a theoretical division: some speakers changed the position of stress within the word, or produced an alternation in the pronunciation (e.g., [ˈtʃʌpteɾ], where the rhotic nucleus is realized as a sequence of V + C). Therefore, the actual position of stress and the actual pronunciation was considered.

Some contexts were excluded from analysis. When a word-final rhotic context is followed by a vowel-initial word, native speakers typically use linking (e.g., "year old" pronounced as [jɪ͡ər‿ə͡ʊld]). Such contexts are irrelevant and thus not considered here, because /r/ would be pronounced in both rhotic and non-rhotic pronunciations. In contrast, non-linking can be included, as there is an option between [jɪ͡ə ʔə͡ʊld] and [jɪ͡ər ʔə͡ʊld]. Finally, foreign proper names were excluded, and mispronunciations were considered only if the target context remained non-prevocalic.

In addition, 35 syllable-initial contexts were selected featuring /r/ as a single onset (i.e., not part of a cluster). These considerations allowed for an investigation of the phonetic realization of the rhotic sound (approximant vs. flap vs. trill).

3.3 Analysis

The occurrence and realization of the rhotic sounds were examined auditorily by the authors with the help of a spectrogram in Praat (Boersma & Weenink, 2014). Tokens were identified as either rhotic or non-rhotic. The former includes a clear presence of a rhotic consonant (as in [ˈɪmpɔːrt]) or a rhotic vowel (mid central vowel with r-colouring spread throughout, as in [ˈfɝːst], [ˈt͡ʃæptɚ]). In the latter case, the /r/-deletion rule was applied, leading to sequences such as [ˈtʃɑːdʒ], [ˈɪmpɔːt], [ˈfɜːst], [ˈt͡ʃæptə] without an audible rhotic element. Furthermore, each rhotic token was coded as "approximant" (standard consonantal realization), "flap" or "trill" (non-standard consonantal realization, e.g., [ˈɪmpoɾt]), or "rhotic vowel" (standard vocalic realization). The very exceptional (and probably unintentional) articulations like uvular [ʁ] or labio-velar [w] were not considered at all. In total, 1724 tokens were extracted in the experimental group (24 speakers) and 141 tokens in the native control group (2 speakers).

The coding of stress was based on the actual realization of stress in each token rather than simply following the citation form of the preceding vowels. For instance, the function words *for* and *or* were coded as unstressed, even though their citation form, which contains the FORCE vowel, could lead one to categorize the vowel as stressed. Similarly, when subjects shifted the position of stress within the word, which often occurred in the less advanced group, the actual position was noted. The position of the rhotic was also labelled according to position within word (final vs. non-final) and following context (whether or not a consonant followed the sound). As a result, three types of contexts are distinguished:

- Absolute final [tʃæptɚ] [wɔːr]
- Preconsonantal final syllable [liːdɚz] [ɪmpɔːrt]
- Preconsonantal non-final syllable [enɚd͡ʒɪ] [pɑːrləmənt]

4 Results

4.1 Overall Degree of Rhoticity

Speakers from the low-proficiency group showed clear rhoticity in their L2 English (92% of tokens were analyzed as rhotic), and the accent preference was irrelevant

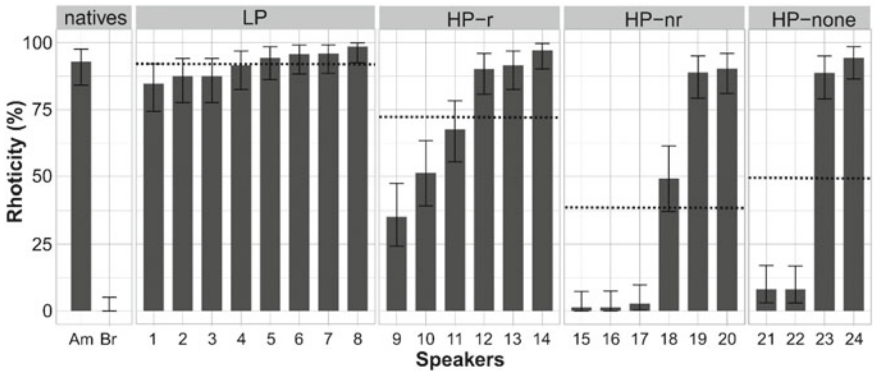

Fig. 1 Rhoticity (mean and 95% confidence intervals) in the native control group and in the Czech-accented group (LP = low proficiency, HP = high proficiency, r vs. nr = striving for a rhotic vs. non-rhotic accent, none = is not striving for a particular accent; dotted lines represent group means)

for rhoticity (and would be difficult to evaluate for only eight speakers in the group anyway). The highly proficient speakers yielded a substantially lower degree of rhoticity on average (54%), which can of course be due to their accent preferences. The preferred model accent indeed seemed to affect rhoticity in the HP speakers considerably: the levels varied from 39% (non-rhotic model) to 72% (rhotic model), with speakers who indicated no accent preference intermediate between the two (49%). These mean values are represented in Fig. 1 with dotted lines (see Sect. 4.2). Statistical evaluation by means of 95% confidence intervals from a binomial test revealed that the rhotic subgroup had significantly higher rhoticity than the other speakers, whereas speakers from the non-rhotic vs. no preference subgroups were not significantly different from each other, as the confidence intervals intersected.

With regard to the linguistic factors examined, only stress affected rhoticity significantly. Stressed syllables were associated with greater rhoticity (70%) than unstressed syllables (64%). In contrast, position within word did not reveal any significant differences in the degree of rhoticity (65–69% on average). Statistical significance was ascertained by the degree of overlap in confidence intervals derived from a binomial test.

4.2 *Individual Variation and Consistency in Rhoticity*

Consistency in rhoticity can be evaluated through comparison to the 50% division line which represents a chance level of rhoticity, that is, when the speaker has no inclination for either type of the rhoticity feature. Therefore, if confidence intervals of a given speaker include the 50% criterion, the speaker can be classified as inconsistent. In case the intervals are sufficiently remote in either direction, the speaker can be classified as predominantly rhotic or non-rhotic. Obviously, there is no level of true consistency apart from 100%, as any other criterion would be arbitrary (e.g., 75%). However, we may still describe the participants in how much they approach the ideal 100% criterion (consult Fig. 1 for the following paragraphs).

As a baseline, two native speakers were recorded. Although the non-rhotic British speaker was truly consistent in her non-rhoticity (0% of rhotic tokens), the rhotic American speaker produced a few non-rhotic tokens, leading to 93% rhoticity. The confidence interval is well above the 50% line, and although the lower bound of the CI reaches 84%, it can be argued that this is still a case of almost absolute consistency.

Learners from the low proficiency group (1–8 in the figure) were remarkably consistent. Four participants even yielded a higher percentage of rhoticity than the native American control speaker, and the others were predominantly rhotic as well. In contrast, the highly proficient learners (9–24) were much more variable, regardless of the model accent preference. With respect to the "rhotic" subgroup, only three speakers (12, 13, 14) seemed to be clearly rhotic, one (10) was truly inconsistent (the CI includes the 50% line), and two (9, 11) were very close to inconsistency, with only a small predominance of non-rhoticity or rhoticity, respectively. What is puzzling is that there is no indication in the questionnaire data about why these three less consistent speakers (9–11) differ from the rest of the subgroup. All three stated they preferred a rhotic accent as both the pronunciation model and the most pleasant-sounding accent. They had none or rhotic native teachers, interacted with none or rhotic native speakers, watched TV series balanced between American and British, and one of them (9) stayed both in the USA and Britain.

The "non-rhotic" subgroup (15–20) also showed a great deal of variation. Three speakers were remarkably non-rhotic (rhotic tokens below 3%), but, surprisingly, two speakers produced predominantly rhotic tokens (around 80%). One of these two learners (19), on the one hand, considered a non-rhotic accent as the best model, but on the other hand, indicated GA as the most pleasant variety, and stayed in the US for more than a month. The other learner (20) had a rhotic native teacher of English. Both watched predominantly American TV. Interestingly, the remaining, inconsistent speaker (18), yielding 49% rhoticity, interacted with both rhotic and non-rhotic speakers, had a rhotic native teacher of English but stayed in Scotland for over a month (and preferred Sottish accent in terms of pleasantness), and watched a balanced set of TV series. Finally, the three consistently non-rhotic learners (15–17) did not have such a rhotic influence in their background.

The third subgroup, which did not state a preference for any accent at the expense of another as a pronunciation model, was seemingly inconsistent as a group (see the mean value represented by the dotted line in Fig. 1, which is close to the 50% line). However, when analyzed at the individual level, two speakers (21, 22) were in fact consistently non-rhotic, whereas the other two (23, 24) were predominantly rhotic. Moreover, all the speakers still indicated a preference for RP (21, 22) or GA (23, 24) in terms of pleasantness, which thus renders their production in alignment with their accent preferences. Their reluctance to designate a model for pronunciation might be a conscious effort not to appear biased towards a specific accent or country, or possibly a truly "neutral" stance. Nevertheless, this does not necessarily mean, and is disproved here, that they do not prefer an accent at a personal level.

4.3 Phonetic Realization of the Rhotic

The second concern with rhoticity is the particular realization of the /r/ phoneme. The learners from the HP group produced in total only four tokens with non-standard, flapped articulation. Therefore, only the LP group was analyzed in this respect.

Alveolar flaps [ɾ] constituted 10% of rhotic tokens in the LP group, the remaining 90% were realized as approximants [ɹ] or rhotic vowels ([ɚ ɝ]). Such number is not surprising, given that the learners estimated their proficiency level to be relatively advanced (mostly B2). Interestingly, alveolar flaps appeared more often in syllabic codas (e.g., "interview" pronounced as [ˈɪnteɾvjuː]) than in nuclei (i.e., substituting the rhotic vowel with the Czech syllabic /r/, as in "minister" pronounced as [ˈmɪnɪstɾ̩]).

Furthermore, the realization of /r/ was examined also in syllable onsets in the LP speakers, that is, in positions where it appears in both rhotic and non-rhotic varieties (e.g., /red/). There were in total 280 such contexts, and alveolar flaps appeared in 16% of the tokens. The difference between the percentage of flaps in the onset and in the rhotic positions was statistically insignificant (confidence intervals overlapped). Trilled articulations [r] did not appear at all. The overall proportion of non-standard articulations roughly corresponds to the findings of Skarnitzl and Rumlová (2019), who reported a value of 15% examining 10 strongly Czech-accented female speakers. It thus seems that Czech speakers do not experience substantial difficulties in acquiring the standard, approximant realization of the rhotic sound, but may have difficulties in producing the unexpected phonotactic distribution of non-rhoticity, or reinterpreting the rhotic vowels as sequences of a vowel plus /r/.

As group averages may be misleading, it is important to examine the individual speakers as well. Figure 2 shows the percentage of alveolar flaps (as opposed to standard realizations) in all the examined contexts together, that is, in both the rhotic targets and the syllable onsets. It is apparent that alveolar flaps appeared predominantly in two speakers only. These two speakers are the oldest persons in the corpus

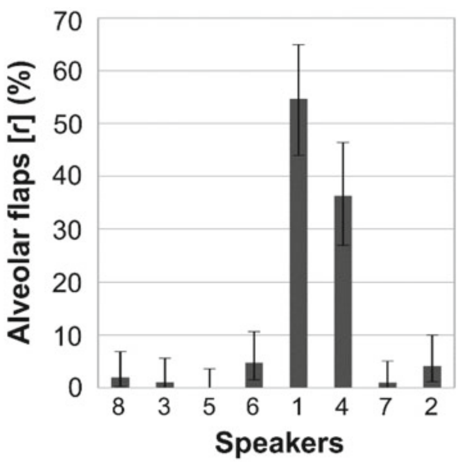

Fig. 2 Proportion of alveolar flaps (as opposed to standard realizations) in all the examined contexts together. Speakers from the low-proficiency group

(59 and 47 years old), and they are also the two latest-onset learners (started learning English at 44 and 26 years old, respectively). Moreover, the oldest and latest-onset speaker was also the one with the lowest estimated level of proficiency (A2). The other speakers in the LP group flapped /r/ rather exceptionally.

5 Discussion and Implications

The results of our research show primarily that rhoticity is indeed typical of the Czech accent in English (see also Kostelná, 2005; Pankovicsová, 2014, for Slovak English). If an L1 Czech speaker wishes to speak in a non-rhotic manner, he or she apparently must be very proficient, with a lot of exposure to native English, and ideally even having spent some considerable time in a non-rhotic English-speaking country. Moreover, it has been shown for Czech secondary-school learners that rhotic tokens are more intelligible and easier to understand than non-rhotic tokens (Kobák, 2017). In the light of this, the pursuit of a non-rhotic accent as the pronunciation model for production in Czech TEFL classrooms, which is currently the case (Brabcová & Skarnitzl, 2018; Jakšič & Šturm, 2017), does not seem to be so reasonable. However, this is not to say that GA should be the model accent, as it obviously possesses other features that Czech learners might find problematic.

On the other hand, the standard realization of /r/ as an approximant or a rhotic vowel does not seem to pose serious difficulties even to intermediate learners, be it in non-prevocalic or prevocalic contexts (see also Skarnitzl & Rumlová, 2019), unless they are older, late-onset learners. This leads us to believe that, in the case of rhoticity, the transfer of the underlying phonemic structures operates as a factor considerably stronger than the transfer of the surface phonetic realization of the rhotic. This cannot be generalized to other features: although Czech learners similarly experience difficulties in acquiring phonemic contrasts such as /e/ vs. /æ/, /n/ vs. /ŋ/, or /θ/ vs. /s/ that are absent in their L1 (Skarnitzl, 2001; see Chap. 11), some other allophonic variation seems to be equally difficult as phonemic structure (Skarnitzl & Šturm, 2016), not to mention prosodic characteristics (Volín et al., 2015).

Curiously enough, when looking at the results globally, it seems as if language proficiency in English decreases the speakers' consistency in rhoticity. The LP learners were all very consistent in their productions, whereas the HP learners appeared to be significantly less consistent—not only when they strived to be non-rhotic (which would not be surprising), but even when they aimed for a rhotic accent. It can be speculated that mere awareness of the existence of non-rhoticity in accents of English lowers consistency, given that the accent-conscious student has to make a decision between the two possibilities, whereas the less advanced learner may simply not know that an <r> in the spelling is not necessarily pronounced in the critical contexts. Unfortunately, no such category as "awareness of accent features" has been tested in the questionnaires.

The social significance of rhoticity cannot be ignored, as famously documented in Labov's experiment from New York City department stores (Labov, 2006 [1966]), although his research was carried out in the 1960s and pertained only to a geographically (not socially) limited community of speakers. The fact that in the nineteenth century the situation was quite opposite in the eastern part of the United States, and non-rhotic pronunciation enjoyed prestige there because of its association with RP, confirms the mutability of standards, but also further proves the prominent position rhoticity has had in the sociolinguistics of English. We can therefore conclude that speaking in a rhotic or non-rhotic manner is an important and defining feature of every English speaker's accent. It is important because it influences a speaker's image in different situations and social and geographical contexts regardless of whether rhoticity is the speaker's deliberate decision or based on cultural and aesthetic preferences, and/or whether the speaker acquired rhoticity/non-rhoticity due to dominant exposure to a particular accent, due to having been taught a particular standard at school, or due to L1–L2 transfer.

One limitation of our study concerns the type of material used to assess the presence or absence of rhoticity. We investigated read speech, which has the advantage of experimental control in that speakers produce identical texts, and a direct comparison of speakers is thus possible. However, rhoticity has been shown to vary with speaking style (Li, 2019; Li & Kabak, 2017; Sundkvist & Gao, 2016), and it can be expected that unconstrained conversation in L2 and reading in L2 will yield somewhat different results. We can speculate that read speech, as a more formal style and with higher level of self-monitoring, will be associated with greater consistency in the more proficient and accent-conscious speakers, especially with regard to non-rhoticity. To complicate the matter further, rhoticity might be affected by the accent and status of the conversational partner (see the research on speech accommodation: Beňuš, 2014; Gallois & Giles, 2015; Pardo, 2013); this would need to be taken into account in conversation, but is not a problem in reading.

As for further and much more ambitious research in this area, it would be very interesting to examine not only conversational speech, but also rhoticity of L1 Czech teachers of English at elementary or secondary schools (experimentally, not by means of questionnaire surveys). The investigation of their L2 speech while teaching would reveal what pronunciation model Czech pupils actually receive, and how consistent the teachers are in producing the model accent. Not too unlikely, it could be some "mildly Czech-accented rhotic (or semi-rhotic) RP" that actually prevails in the teachers' speech.

6 Conclusion

Rhoticity is a distinct and readily recognizable feature of every accent of English. The description of Czech learners' tendencies in this specific aspect might therefore contribute to creating a more comprehensive image of contemporary Czech foreign accent in English. The results of our study examining the speech of 24 participants differing in proficiency level revealed that less advanced Czech speakers of English tended to be consistently rhotic, and generally able to pronounce the rhotic contexts in a standard manner. In contrast, the group of advanced learners—university students of English—showed substantial variation in the consistency of rhoticity. First, the extent of rhoticity to a large degree corresponded to their accent preferences in terms of model accent and accent pleasantness. Second, there was further variation within these groups: predominantly rhotic individuals could still regard a prestigious non-rhotic accent as a model for pronunciation, and vice versa. Importantly, a number of background variables characterizing the amount of contact with speakers of rhotic versus non-rhotic varieties was usually—but not always—helpful in explaining these anomalies.

Acknowledgements The work was supported by the European Regional Development Fund-Project 'Creativity and Adaptability as Conditions of the Success of Europe in an Interrelated World' (No. CZ.02.1.01/0.0/0.0/16_019/0000734).

Appendix

The questionnaire filled in by the speakers (translation into English).
Sex ___ Age ___ Region of origin ___ Mother tongue ___
How many years have you studied English? ___
Do you study English as a field (study programme) at university? ___
At which level of English according to CEFR do you estimate yourself to be? (A1, A2, B1, B2, C1, C2)
What other languages do you speak at the C1/C2 level? ___
Which of the following accents of English is the most pleasant to you? (British [standard, RP], American [standard, GA], Australian, Canadian, Scottish, other ___)
Do you aim to emulate consciously any of these accents? If so, which one? ___
If you had native teachers of English at school, what nationality were they? ___
If you speak regularly with native speakers of English, what nationality are they? ___
In which English-speaking countries have you been for at least 1 month? ___
Do you watch / have you watched any of the following series in the original language? (Friends / Game of Thrones / Red Dwarf / The Simpsons / Sherlock / Breaking Bad / IT Crowd / Black Books / Dr House / How I Met Your Mother / The Big Bang Theory / Doctor Who / other ___)

References

Asprey, E. (2007). Investigating residual rhoticity in a non-rhotic accent. *Leeds Working Papers in Linguistics and Phonetics*, *12*, 78–101.

Becker, K. (2014). (r) we there yet? The change to rhoticity in New York City English. *Language Variation and Change, 26*(2), 141–168. https://doi.org/10.1017/S0954394514000064

Beňuš, Š. (2014). Social aspects of entrainment in spoken interaction. *Cognitive Computation, 6*(4), 802–813. https://doi.org/10.1007/s12559-014-9261-4

Boersma, P., & Weenink, D. (2014). *Praat: Doing phonetics by computer* (Version 5.4) [Computer software]. http://www.praat.org

Brabcová, K., & Skarnitzl, R. (2018). Foreign or native-like? The attitudes of Czech EFL learners towards accents of English and their use as pronunciation models. *Studie z Aplikované Lingvistiky, 9*(1), 38–50.

Brown, A. (2015). Syllable structure. In M. Reed & J. M. Levis, (Eds.), *The handbook of English pronunciation* (pp. 85–105). Wiley. https://doi.org/10.1002/9781118346952.ch5

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

Chiba, R., Matsuura, H., & Yamamoto, A. (1995). Japanese attitudes toward English accents. *World Englishes, 14*(1), 77–86. https://doi.org/10.1111/j.1467-971X.1995.tb00341.x

Dalton-Puffe, C., Kaltenböck, G., & Smit, U. (1997). Learner attitudes and L2 pronunciation in Austria. *World Englishes, 16*(1), 115–128. https://doi.org/10.1111/1467-971X.00052

Deterding, D. (2007). *Singapore English*. Edinburgh University Press. https://doi.org/10.3366/edinburgh/9780748625444.001.0001

Gallois, C., & Giles, H. (2015). Communication accommodation theory. In *The international encyclopedia of language and social interaction* (pp. 1–18). Wiley. https://doi.org/10.1002/9781118611463.wbielsi066

Hansen Edwards, J. G. (2016). Accent preferences and the use of American English features in Hong Kong: A preliminary study. *Asian Englishes, 18*(3), 197–215. https://doi.org/10.1080/13488678.2016.1225482

Hiang, T. C., & Gupta, A. (1992). Post-vocalic /r/ in Singapore English. *York Papers in Linguistics, 16*, 139–152.

Jakšič, J., & Šturm, P. (2017). Accents of English at Czech schools: Students' attitudes and recognition skills. *Research in Language, 15*(4), 353–369. https://doi.org/10.1515/rela-2017-0020

Kang, H. (2013). Internal and external constraints on rhoticity in Korean English. *The Sociolinguistic Journal of Korea*, *21*(2), 1–27. https://doi.org/10.14353/sjk.2013.21.2.1

Kobák, A. (2017). *(Non)rhoticity in English pronunciation teaching* [Unpublished master's thesis]. Charles University.

Kostelná, K. (2005). *Rhoticity in Czech students and graduates' English* [Unpublished BA thesis]. Charles University.

Kuecker, K., Lockenvitz, S., & Müller, N. (2015). Amount of rhoticity in schwar and in vowel+/r/ in American English. *Clinical Linguistics & Phonetics, 29*(8–10), 1–7. https://doi.org/10.3109/02699206.2015.1044674

Labov, W. (1966/2006). *The social stratification of English in New York City*. Cambridge University Press. https://doi.org/10.1017/CBO9780511618208

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Wiley-Blackwell.

Ladegaard, H., & Sachdev, I. (2006). 'I Like the Americans… But I certainly don't aim for an American accent': Language attitudes, vitality and foreign language learning in Denmark. *Journal of Multilingual and Multicultural Development, 27*(2), 91–108. https://doi.org/10.1080/01434630608668542

Li, J. (2019). (r) we Americanised? The emerging rhoticity features in China English. *English Today, 35*(1), 28–35. https://doi.org/10.1017/S0266078418000111

Li, Z., & Kabak, B. (2017). Rhoticity in Chinese English: An experimental investigation on the realization of the variant (r) in an Expanding Circle variety. *Alicante Journal of English Studies, 30*, 61–92. https://doi.org/10.14198/raei.2017.30.03

Marsden, S. (2017). Are New Zealanders "rhotic"? The dynamics of rhoticity in New Zealand's small towns. *English World-Wide, 38*(3), 275–304. https://doi.org/10.1075/eww.38.3.02mar

Mompeán, J. A. (2008). Consumers' preferences and the choice of English pronunciation models. In R. Monroy & A. Sánchez Pérez (Eds.), *25 years of applied linguistics in Spain: Milestones and challenges* (pp. 959–964). University of Murcia.

Pankovicsová, A. (2014). *Slovak English: Rhotic or non-rhotic?* [Unpublished master's thesis]. Charles University.

Pardo, J. S. (2013). Measuring phonetic convergence in speech production. *Frontiers in Psychology, 4*, Article 559. https://doi.org/10.3389/fpsyg.2013.00559

Piercy, C. (2012). A transatlantic cross-dialectal comparison of non-prevocalic /r/. *University of Pennsylvania Working Papers in Linguistics, 18*(2), 77–86.

Reed, M., & Levis, J. (Eds.). (2015). *The handbook of English pronunciation.* Wiley. https://doi.org/10.1002/9781118346952

Schneider, E., Burridge, K., Kortmann, B., Mesthrie, R., & Upton, C. (Eds.). (2004). *A handbook of varieties of English: Volume 1: Phonology.* De Gruyter Mouton. https://doi.org/10.1515/9783110197181

Šimáčková, Š. (2003). "Engela's Eshes": Cross-linguistic perception and production of English [æ] and [ε] by Czech EFL learners trained in phonetics. In M.-J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 2293–2296). Causal Productions.

Skarnitzl, R. (2001). *Teaching and learning the English dental fricatives in the Czech environment* [Unpublished master's thesis]. Charles University.

Skarnitzl, R., & Rumlová, J. (2019). Phonetic aspects of strongly-accented Czech speakers of English. *Acta Universitatis Carolinae – Philologica*, *2019*(2), 109–128. https://doi.org/10.14712/24646830.2019.21

Skarnitzl, R., & Šturm, P. (2016). Pre-fortis shortening in Czech English: A production and reaction-time study. *Research in Language, 14*(1), 1–14. https://doi.org/10.1515/rela-2016-0005

Skarnitzl, R., & Šturm, P. (2017). Voicing assimilation in Czech and Slovak speakers of English: Interactions of segmental context, language and strength of foreign accent. *Language and Speech, 60*(3), 427–453. https://doi.org/10.1177/0023830916654509

Stuart-Smith, J., Lawson, E., & Scobbie, J. M. (2014). Derhoticisation in Scottish English: A sociophonetic journey. In C. Celata & S. Calamai (Eds.), *Advances in sociophonetics* (pp. 57–94). John Benjamins. https://doi.org/10.1075/silv.15.03stu

Šturm, P., & Skarnitzl, R. (2011). The open front vowel /æ/ in the production and perception of Czech students of English. In P. Cosi, R. De Mori, G. Di Fabbrizio, & R. Pieraccini (Eds.), *Proceedings of 12th Annual Conference of the International Speech Communication Association 2011 (INTERSPEECH 2011)* (pp. 1161–1164). ISCA. https://doi.org/10.21437/Interspeech.2011-344

Sundkvist, P., & Gao, M. (2016). Rhoticity in Yunnan English. *World Englishes*, *35*(1), 42–59. http://dx.doi.org/10.1111/weng.12172

Volín, J., & Johaníková, T. (2018). Weak structural words in British and Czech English. In J. Volín & R. Skarnitzl (Eds.), *The pronunciation of English by speakers of other languages* (pp. 181–195). Cambridge Scholars.

Volín, J., Poesová, K., & Weingartová, L. (2015). Speech melody properties in English, Czech and Czech English: Reference and interference. *Research in Language, 13*(1), 107–123. https://doi.org/10.1515/rela-2015-0018

Waniek-Klimczak, E., & Klimczak, K. (2008). Target in speech development: Learners' views. In K. Dziubalska-Kołaczyk & J. Przedlacka (Eds.), *English pronunciation models: A changing scene* (pp. 229–249). Peter Lang.

Wells, J. C. (1982). *Accents of English.* Cambridge University Press. https://doi.org/10.1017/CBO9780511611759

Young, T. J., & Walsh, S. (2010). Which English? Whose English? An investigation of "non-native" teachers' beliefs about target varieties. *Language, Culture and Curriculum, 23*(2), 123–137. https://doi.org/10.1080/07908311003797627

Ondřej Fischer studies the English language at Charles University, and musical composition at Jaroslav Ježek Conservatory in Prague, Czech Republic. His main areas of interest are English language teaching and language variation and change. In addition, he works as a presenter for a classical radio station and as a teacher of English for a faculty language school.

Pavel Šturm is Assistant Professor at the Faculty of Arts, Charles University, Prague, Czech Republic. His research interests lie mainly in the domain of second language pronunciation (especially L2 English), focusing on segmental aspects in terms of production and on foreign accent evaluation and perception. He also examines the issues of syllabification and phonotactics in the Czech language.

Phonetic Training

High-Variability Phonetic Training Under Different Conditions: Individual Differences in Auditory Attention Control

Ingrid Mora-Plaza, Mireia Ortega, and Joan C. Mora

Abstract Cognitive attention control guides auditory processes during speech processing but its contribution to L2 speech learning remains under-researched. This study examined the interaction between individual differences in auditory selective attention (ASA) and attention switching (ASW), and the effectiveness of high-variability phonetic training (HVPT) administered under different stimuli and presentation conditions to improve L2 learners' sensitivity to an L2 vowel contrast and its lexical encoding. Catalan-Spanish bilingual adult learners of English ($N = 102$) were randomly assigned to eight HVPT groups and trained in four 35-min sessions on the perception and production of English /æ/ and /ʌ/ through identification, discrimination, and immediate repetition tasks. Learners' gains were assessed through ABX discrimination and delayed word repetition tasks. Lexical encoding was tested through lexical decision and delayed sentence repetition tasks. We measured ASA through a single-talker competition paradigm and ASW through a novel speech-based version of the alternating-runs task-switching paradigm. Results showed that ASA was often related to pre-test (T1) and post-test (T2) perception accuracy but unrelated to either production accuracy or T1-T2 perceptual and production gains. However, ASW was related to /æ/ and /ʌ/ perception and production gains, but this varied as a function of stimuli type and presentation condition.

Keywords Attention control · HVPT · L2 vowel perception · L2 vowel production · Lexical encoding

I. Mora-Plaza (✉) · M. Ortega · J. C. Mora
University of Barcelona, Barcelona, Spain
e-mail: imoraplaza@ub.edu

M. Ortega
e-mail: m.ortega@ub.edu

J. C. Mora
e-mail: mora@ub.edu

V. G. Sardegna and A. Jarosz (eds.), *Theoretical and Practical Developments in English Speech Assessment, Research, and Training*, Second Language Learning and Teaching,
https://doi.org/10.1007/978-3-030-98218-8_14

1 Introduction

Although cases of exceptional L2 phonological acquisition have been attested in the Second Language Acquisition (SLA) literature (Moyer, 2014), most L2 learners struggle with L2 pronunciation, especially in instructed foreign language learning contexts where opportunities for L2 exposure and use are generally scarce. Experience-related factors that have been shown to explain inter-learner variation in L2 pronunciation learning in immersion settings, such as amount of L1 and L2 use, age of onset of L2 learning, length of residence, L2 input quantity and quality, among others (Flege, 2009; Munro & Bohn, 2007), have been shown to play a modest role in instructed SLA (Cebrian, 2006). However, both in immersion and instructed foreign language settings, individual differences in L2 phonological attainment cannot fully be accounted for by experience-related factors alone. Socio-psychological factors such as motivation, anxiety, or willingness to communicate (Kormos, 2017) as well as cognitive and aptitude-related factors such as auditory processing (Saito et al., 2020), working memory, inhibition, and attention (Darcy et al., 2014; Ghaffarvand Mokari & Werner, 2019; Lev-Ari & Peperkamp, 2014) also play a role.

Given the myriad of factors affecting L2 phonological acquisition over time and their interaction with L2 learners' individual differences, identifying, isolating, and quantifying the independent contribution of specific cognitive variables (e.g., attention control) to L2 speech learning becomes a challenging research objective. Two features make laboratory-based phonetic training an optimal testing ground: (a) variability in the extent to which learners benefit from it, and (b) full control over the type and amount of input learners receive (Golestani & Zatorre, 2009). Under such conditions, gains in perception and production can be directly related to independent measures of cognitive control.

The current study sets out to explore the role of cognitive attention control in L2 speech learning by examining the interaction between individual differences in auditory selective attention (ASA) and auditory attention switching (ASW) skills, and the effectiveness of high-variability phonetic training (HVPT) administered under different stimuli and presentation conditions. We focused on L1-Spanish/Catalan advanced learners' perception and production of English /æ/-/ʌ/ and its lexical encoding.

2 Literature Review

2.1 Phonetic Training

Most previous phonetic training research has used either perception (Bradlow, 2008) or production training methods (Kartushina et al., 2015). In perception, identification training has generally been found to lead to larger gains than discrimination training (Carlet & Cebrian, 2019), but few studies have combined discrimination and identification training (Shinohara & Iverson, 2018) or perception and production training tasks in a HVPT paradigm (Wong, 2013). Additionally, phonetically-oriented training with nonwords has been shown to lead to larger gains than training with words because non-lexical materials allow learners to focus on the phonetic properties of the training stimuli while avoiding interference from lexically misrepresented phonetic forms (Ortega et al., 2021; Thomson & Derwing, 2016). Auditory attention control skills may potentially have a differential impact on training gains under phonetically- and lexically-oriented conditions. For example, as hypothesized in the current study, ASA may play a fundamental role in phonetically-oriented training, allowing learners to more easily extract the relevant phonetic properties that distinguish the target vowels /æ/ and /ʌ/. On the other hand, ASW, which involves inhibiting phonetic dimensions not under focus, may be more relevant in lexically-oriented training, where learners are trained on phono-lexical forms that are not likely to match their own representations.

Some training conditions have been shown to lead to greater gains. For instance, the presence of noise during training has been proved to have the effect of degrading the intelligibility of the speech signal (Mattys et al., 2012), but at the same time, it may help learners focus their attention on the more robust phonetic properties distinguishing the target contrast (Cooke & García-Lecumberri, 2018), and in production training it may lead to hyper-articulated speech (Hazan & Baker, 2011), which may enhance learners' ability to distinguish the target vowels in production. Audiovisual phonetic training has been shown to be superior to auditory-only training in training L2 sound contrasts (Hazan et al., 2005), and visual feedback has proved particularly effective in training the production of L2 vowels (Kartushina et al., 2015).

2.2 Attention Control in L2 Speech Learning

Attention control is implicated in speech processing and language comprehension and production (Miyake & Friedman, 2012) and in second language acquisition (Segalowitz & Frenkiel-Fishman, 2005). Both ASW and ASA skills allow listeners to selectively attend to specific acoustic dimensions during speech processing and to focus their processing resources on the auditory information that is relevant for language decoding processes to work efficiently (Astheimer et al., 2016). ASA skills, additionally, allow listeners to selectively attend to a single acoustic dimension or feature during speech processing, thus facilitating perceptual learning and the processing of L2 phonological contrasts (Ou et al., 2015). Phonetic training is effective in training learners to attend to speech dimensions and L2-specific acoustic cues not attended to in their native language (Iverson et al., 2005), suggesting that attention control skills may be an important source of individual differences in L2 phonetic training.

Research on the role of attention control in L2 phonetic training is scarce and has produced mixed results. For example, Kim and Hazan (2010) found ASW skills to be related to training gains in naïve L1-English speakers trained to perceive a novel Korean stop voicing contrast. Mora and Mora-Plaza (2019) trained L1-Spanish learners in the perception and production of two L2-English vowel contrasts (/æ/-/ʌ/ and /iː/-/ɪ/). They found ASA to explain gains in the perception of one contrast (/æ/-/ʌ/), but not the other (/iː/-/ɪ/) and ASW was related to accuracy of performance in perceptual discrimination tasks, but unrelated to perception training gains. In the same line, Ghaffarvand Mokari and Werner (2019) found attention control to be unrelated to training gains for L1-Azerbaijani learners of English.

3 The Study

The main aim of this study is to examine the extent to which individual differences in auditory attention control can explain inter-learner variability in training gains for a challenging L2 vowel contrast. We chose the /æ/-/ʌ/ contrast because it is a difficult L2 contrast for L1-Spanish and L1 Spanish-Catalan bilingual learners of English alike (Rallo-Fabra & Romero, 2012), as both English vowels are perceptually mapped onto a single L1 low central vowel category /a/ in Spanish and Catalan, although /æ/ is a slightly better perceptual match for Spanish and Catalan /a/ than English /ʌ/ (Cebrian, 2019; Cebrian et al., 2011). To maximize potential training gains, we used a comprehensive HVPT paradigm that included two perception and one production task in every training session (see Sect. 4.3.1). Finally, to investigate potential interactions of cognitive attention control with training conditions requiring differential use of attentional resources, we trained learners with nonwords or with words. We also trained them with or without noise, and with or without visual monitoring. Based on Cooke and García-Lecumberri (2018), we expected learners with stronger auditory attention control skills to be better able to focus attention on the target vowels during stimuli repetition in the presence of masking noise. Additionally, we assessed the potential benefits of visual monitoring (watching one's own lips) during production training (with and without noise). Based on Hardison (2018), strong auditory attention control should allow learners to benefit from visual cues enhanced through the presence of masking noise.

The following research questions guided our investigation:

1. Does HVPT improve the perception and production of /æ/ and /ʌ/?
2. Does HVPT improve the lexical encoding of the /æ/-/ʌ/ contrast?
3. Do individual differences in auditory attention control explain variance in training gains?
4. To what extent does auditory attention control interact with training conditions to explain training gains?

4 Methods

4.1 Participants

The participants were 116 Spanish-Catalan bilingual undergraduate learners of English (see Table 1 for demographics) randomly assigned to one of eight different experimental training groups ($N = 102$) or to an untrained control group ($N = 14$; Table 2). One-way ANOVAs with *Training Group* as the independent variable confirmed that the experimental groups were comparable in L2 proficiency, $F(7,93) = 0.688$, $p = 0.681$, and L2 vocabulary size, $F(7,88) = 0.436$, $p = 0.877$. All participants reported having no speech or hearing pathologies.

4.2 Materials

The testing and training word and nonword stimuli contained the target vowels /æ/ and /ʌ/ as produced by six southern British English speakers (3 females, 3 males). They were elicited in carrier phrases (*I say X, I say X again*), recorded in a soundproof booth, excised, and normalized for amplitude in Praat (Boersma & Weenink, 2020). Four voices were used in the training and two of them (1 female, 1 male) were used for the testing stimuli only. Training stimuli were high-variability monosyllabic CVC nonword (8) and word (8) minimal pairs with the target vowels in eight different phonetic environments (e.g., *chang* /ʧæŋ/, *chung* /ʧʌŋ/, *mad* /mæd/, *mud* /mʌd/). Testing stimuli consisted of 12 monosyllabic CVC nonword minimal pairs (6 trained, 6 untrained) and 18 monosyllabic CVC word minimal pairs (6 trained, 12 untrained), plus 16 words which were presented in isolation and in the context of a sentence.

4.3 Procedure

Participants completed a language background questionnaire, and then they were trained individually in four 35-min sessions in a quiet lab, twice per week for two consecutive weeks (see training tasks in Sect. 4.3.1) and pre-and post-tested immediately before and after the training (see testing tasks in Sect. 4.3.2). Participants' cognitive attention control was measured in Session 2 (see cognitive control attention tasks in Sect. 4.3.3). Finally, participants' L2 proficiency was assessed in Session 3 via an elicited imitation (EI) test (Ortega et al., 2002) consisting of 30 sentences varying in length (7–19 syllables) and grammatical complexity. Participants had to repeat the sentences from memory after a 2000 ms delay. They also completed a yes/no vocabulary knowledge test (X/Y Lex; Meara & Miralpeix, 2006) that provided a measure of receptive vocabulary size (0–10,000 words). Figure 1 displays the distribution of training and testing tasks, and the attention control and L2 proficiency tasks.

Table 1 Participants' demographics

Measure	G1		G2		G3		G4		G5	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Age at testing (years)	24.7	9.8	23.9	10.1	21.2	2.3	21.7	4.6	22.2	7.4
L2 learning onset age (years)	7.2	2.9	7.3	4.1	6.1	1.9	5.2	2.2	6.1	2
Spoken L2 input	18.1	11	13.8	7.5	10.9	3.6	11.6	5.9	8.1	4.9
Spoken L2 output	7.1	5.3	5.7	3.9	5.9	3.3	5.9	5.6	2.9	2.6
L2 use[a]	9.8	3.8	14.2	6.5	12	7.4	12.2	6.9	13.9	9.7
Vocab. size (0–10,000 words)[b]	6431	1243	6491	1442	6215	1124	6579	1200	6277	1310
L2 proficiency (0–120 points)[c]	93.3	11.6	89.3	21.4	94.9	12.1	97.1	10.9	93.6	18.9
Self-estimated proficiency (1 = very poor—9 = native-like)[d]	6.7	0.8	6	1.5	7.2	1.0	7	0.7	6.8	1.2

Measure	G6		G7		G8		Control	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Age at testing (years)	20.4	1.1	22.6	10.2	22.3	7.4	26.7	7.1
L2 learning onset age (years)	4.5	2.6	5.6	1.8	5.8	1.9	7.7	6.5
Spoken L2 input	16.3	9	11	5.9	15.6	8.8	12.6	9.7
Spoken L2 output	5.8	3.7	4.9	2.6	7.6	4.7	4.9	4.3
L2 use[a]	13	4	17.1	10.9	10.5	4.4	12.5	6.2
Vocab. size (0–10,000 words)[b]	5891	1310	6408	1196	5923	1141	–	–
L2 proficiency (0–120 points)[c]	94.5	15.9	100.2	15.6	89.6	9.1	–	–
Self-estimated proficiency (1 = very poor—9 = native-like)[d]	6.3	2.3	6.9	1.4	6.4	1	6.4	2

[a] L2 use with native and non-native speakers in hours per week
[b] Obtained through the X/Y Lex test (Meara & Miralpeix, 2006)
[c] Obtained through the Elicited Imitation task (Ortega et al., 2002)
[d] Averaged self-estimated ability to speak spontaneously, understand, read, write and pronounce English

Table 2 Participant groups and training conditions

		Production training conditions		
Group	*N*	Stimuli Type	Monitoring	Listening
G1	13	Nonwords (NW)	Visual (V)	Noise
G2	11			Silence
G3	13		Auditory-only (A)	Noise
G4	14			Silence
G5	13	Words (W)	Visual	Noise
G6	12			Silence
G7	14		Auditory-only	Noise
G8	12			Silence

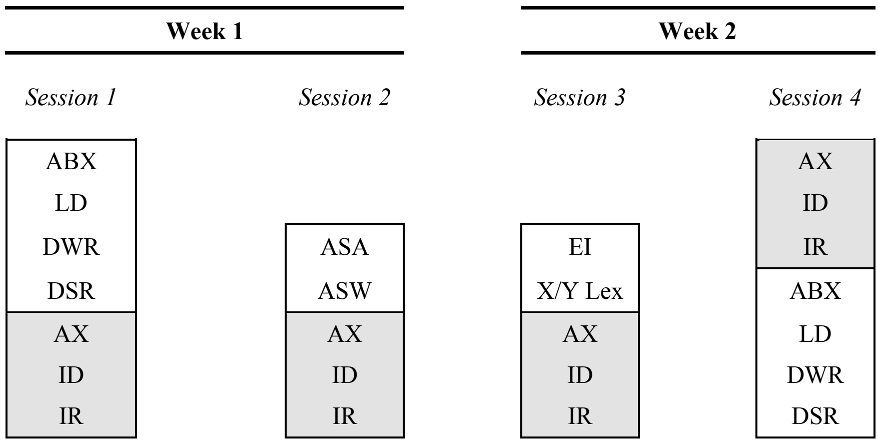

Fig. 1 Distribution of testing and training tasks (shading identifies training tasks)

4.3.1 Phonetic Training

The eight training groups differed in the type of stimuli they were trained on (nonwords or words) and the conditions in which they were administered during production training (with or without noise and/or visual monitoring) (Table 2).

In each of the four training sessions learners were trained perceptually through AX discrimination and identification tasks, and productively through an immediate repetition task (in this order, see Fig. 1).

- AX Discrimination (AX): Participants heard two stimuli (ISI = 500 ms) and decided (as fast and accurately as they could) whether the second vowel in the stimuli (X) contained the same English vowel as the first (same) or not (different). Participants responded to four practice trials and 96 test trials in every session (96 $\times$ 4 = 384 trials) to which they received feedback on accuracy and response latency in milliseconds. The task contained the same number of same (AA, BB) and different trials (AB, BA), and combined a female and a male voice within trials. This perception task was included as a complement to identification training (Shinohara & Iverson, 2018) to increase learners' sensitivity to the primary acoustic cues qualitatively distinguishing /æ/ from /ʌ/ (1st and 2nd formant frequencies) and to improve their pre-categorical processing.

- Identification (ID): Participants heard one stimulus and identified (as fast and accurately as they could) whether it contained the vowel in the word *cap* or in the word *cup* by pressing a designated key on the keyboard matching the corresponding word, which appeared (together with its phonetic transcription and a picture representing it) on the bottom left or right side of the screen. Participants responded to four practice trials and 32 test trials in every session ($32 \times 4 = 128$ trials) and received feedback as in the AX task. This perception task was intended to improve category representations for /æ/ and /ʌ/ and their categorical processing in order to enhance generalization across contexts and talkers (Sadakata & McQueen, 2013).
- Immediate Repetition (IR): Participants heard the same stimuli as those in the ID task and were asked to repeat them twice as accurately as they could focusing on the vowel sound. They heard one stimulus, had 2000 ms to repeat it, then they heard it again, and had 2000 ms more to repeat it again. This procedure allowed learners to monitor their own productions. Participants responded to four practice trials and 32 test trials in every session ($32 \times 4 = 128$ trials). The training conditions for this task varied depending on the experimental group (Table 2) in terms of stimuli type (nonwords vs. words) and presentation condition (with or without noise and visual monitoring). This production task was included to allow participants to implement articulatory changes in the production of the contrast as they learned to perceptually discern /æ/ from /ʌ/. In this task, masking noise was included to enhance the production of clear speech in the auditory-only condition and to enhance attention to articulatory visual cues in the visual monitoring condition.

4.3.2 Testing

Vowel perception and production was pre- and post-tested through an ABX discrimination task and a delayed word repetition (DWR) task, respectively. The lexical encoding of the target vowel contrast was pre- and post-tested in perception and production through a Lexical Decision (LD) task and a delayed sentence repetition task (DSR), respectively (see Fig. 1).

- ABX Discrimination (ABX): Participants heard three stimuli in a row (ISI = 500 ms) and decided within 2500 ms (as fast and accurately as they could) whether the third one (X) contained the same vowel as the first (A) or the second (B) stimulus. Participants responded to a total of 136 trials: 30 test trials in four orders (ABA, ABB, BAB, BAA) = 120; and 8 control trials (/æ/-/iː/, /ʌ/-/iː/).
- Delayed Word Repetition (DWR): Participants repeated the words and nonwords they heard after a tone signal presented 1500 ms after stimulus onset. This delayed presentation procedure avoided repetition from sensory memory and ensured the

elicited stimuli reflected participants' vowel representations. To test for generalization effects, the testing stimuli contained trained and untrained words and nonwords in two different untrained voices (1 female, 1 male).

- Lexical Decision (LD): Participants heard the stimuli in a novel female speaker's voice and decided whether they were real or fake English words. Out of the 56 trials in the test, half were fillers (e.g., *lake*), and the other half were 14 word (e.g., *map, sun*) and 14 nonword (e.g., *mup, san*) test trials with an equal number of /æ/ and /ʌ/ items (half words and half nonwords). We used the proportion of correctly identified nonwords (e.g., *mup* or *san*) as a measure of perceptual sensitivity to the target contrast in a lexical context.
- Delayed Sentence Repetition (DSR): Participants silently read a sentence appearing on the screen (e.g., *He looked at the map to find his way*) targeting an /æ/ or /ʌ/ word (e.g., *map*), then they heard the sentence without reading it, and then waited 1500 ms for a tone signal to repeat it from memory. Sixteen sentences in untrained voices (1 female, 1 male) were repeated twice. Vowels elicited this way were deemed to reflect their corresponding category representations as encoded in the learners' mental lexicon.

4.3.3 Cognitive Attention Control

In Session 2, participants carried out two cognitive attention control tasks (see Fig. 1).

- Auditory Selective Attention (ASA) (Humes et al., 2006): This task consisted of 64 trials of pairs of English sentences (target vs. competitor). The two sentences in a pair were always different, one spoken by a female voice and the other by a male voice and were presented simultaneously through both ears. In every trial, a word signal (e.g., CHARLIE) appeared on the screen cueing the voice participants had to pay attention to in the sentences they would hear simultaneously (e.g., "*Ready Charlie go to blue six now*" + "*Ready Tiger go to red four now*"). Participants identified 1 of 4 colours and 1 of 8 digits visually presented on the screen (e.g., *blue* and *six* for the word signal CHARLIE). In this way, one of the voices and spoken sentences had to be attended to in order to correctly identify the colour and digit while the other was inhibited. Scores could range 0–128, one point for correctly identified colour and digit.
- Auditory Attention Switching (ASW): This task required participants to attend to either the duration (quantity) or the voice (quality) of L1 Catalan vowels (Safronova & Mora, 2013). Tokens of seven isolated Catalan vowels /i e ɛ a ɔ o u/ produced by a male and a female speaker were manipulated in Praat (Boersma & Weenink, 2020) to create short (200 ms) and long (500 ms) versions of the seven vowels. Eight identical copies of each stimulus ($28 \times 8 = 224$ trials) were randomly presented to participants over headphones for categorization as either long/short or male/female. The location of a speaker icon appearing predictably in clockwise fashion together with each auditory stimulus in one of four boxes cued the dimension to be attended to: long/short when appearing in one of the two top boxes, male/female when appearing in one of two bottom boxes.

Within-dimension (repeat trials) response times (RTs) were expected to be shorter than across-dimension (switch trials) RTs. A shorter switch-cost RT score (switch RT minus repeat RT) reflected stronger ASW skills.

The perception and production tasks and the ASW test were administered in *DmDx* (Forster & Forster, 2003), the ASA test in *Inquisit* (Draine, 1999). Participants' productions were recorded at a sampling frequency of 44.1 kHz on Marantz PMD-661 digital recorders with an external Shure SM58 voice microphone.

4.4 Data Analysis

For the ABX and LD tasks, we obtained accuracy and RT scores. RT scores included correct responses only and were screened to exclude RTs 2.5 *SD*s below or above each subject's mean. For the DWR and DSR tasks, we computed vowel production accuracy scores as the spectral distance between participants' vowel production and the average of the same vowels in the same items as produced by the six native speakers whose voices were used in the testing. Vowel frequency measures ($f0$, F1, F2) were extracted in Praat (Boersma & Weenink, 2020) from a 10-ms window centred at the midpoint of the steady-state portion of the target vowels. Extreme values above or below 3 *SDs* from each participant's mean were replaced with the mean value for that vowel in the same testing time. To minimize age, gender, and vocal tract size differences, frequency values in Hertz (Hz) were converted to Bark (B), and then a Bark-distance normalization procedure was used to provide speaker-independent estimates of vowel quality. The difference in Bark between F1 and $f0$ (B1-B0) estimated vowel height, whereas the difference between F2 and F1 (B2-B1) estimated vowel frontness (Bohn & Flege, 1990).

Scores from all tasks were fitted to Generalized Linear Mixed Models (GLMMs) in SPSS 25, with *Testing Time* (T1, T2), *Group* (G1-G9), and *Vowel* (/æ/, /ʌ/) as fixed effects, and *Subject* and *Item* as random factors. To assess the relationship between attention control and training gains, we aggregated the scores by subject and ran *Pearson-r* correlations.

5 Results

First, we present the results by group in terms of the effects of training on participants' sensitivity to the contrast (ABX and DWR) and its lexical encoding (LD and DSR). Second, we report the results on the relationship between cognitive attention control (ASA and ASW) and perception and production training gains and performance.

5.1 *Training Effects on /æ/ and /ʌ/ Perception and Production*

In general, vowel perception and production accuracy (ABX and DWR) improved for all groups (Table 3), and the lexical encoding (LD and DSR) of the contrast did, too, but to a lesser extent, except for the control group (G9), who did not show improvement in any testing task.

For ABX accuracy, the GLMM revealed a significant main effect of *Testing Time,* $F(1{,}28524) = 203.352$, $p < 0.001$, and *Vowel,* $F(1{,}28524) = 254.430$, $p < 0.001$, and a significant *Group* × *Testing Time* × *Vowel* interaction, $F(8{,}28524) = 2.787$, $p = 0.004$. This interaction arose because only G3 (NW + A + noise), G4 (NW + A + silence), G6 (W + V + silence), and G7 (W + A + noise) significantly improved on both vowels (see Tables 2 and 3). No other main effects or interactions reached significance.

Table 3 Descriptive statistics for ABX (proportion of correct responses), LD (proportion of correctly identified nonwords), DWR and DSR (spectral distances in Bark between learners' and native speakers' productions), by vowel and group. Shading indicates improvement *(M = mean, SD = standard deviation)*

	ABX								**DWR**							
	/æ/				**/ʌ/**				**/æ/**				**/ʌ/**			
	T1		**T2**		**T1**		**T2**		**T1**		**T2**		**T1**		**T2**	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
G1	.78	.41	.83	.37	.70	.45	.74	.44	1.99	1.56	1.77	1.38	2.16	1.52	1.96	1.42
G2	.80	.40	.83	.37	.70	.46	.78	.41	1.38	0.89	1.31	0.95	1.48	0.96	1.52	0.98
G3	.71	.45	.81	.39	.68	.46	.73	.44	1.58	1.30	1.33	0.97	1.80	1.30	1.80	1.36
G4	.75	.43	.83	.37	.65	.47	.75	.43	1.60	1.38	1.32	1.04	1.92	1.46	1.66	1.28
G5	.77	.42	.86	.34	.71	.45	.75	.44	1.47	1.31	1.58	1.37	1.72	1.50	1.70	1.40
G6	.78	.41	.85	.35	.70	.46	.81	.39	1.39	1.12	1.18	0.84	1.49	1.06	1.34	0.89
G7	.77	.42	.81	.38	.67	.47	.79	.41	1.35	1.16	1.30	1.10	1.48	1.13	1.46	1.13
G8	.80	.40	.83	.37	.67	.47	.78	.41	1.28	1.02	1.30	0.96	1.49	1.00	1.44	1.02
G9	.77	.42	.77	.37	.64	.46	.65	.45	1.24	0.94	1.38	1.17	1.45	1.04	1.57	1.13

	LD								**DSR**							
	/æ/				**/ʌ/**				**/æ/**				**/ʌ/**			
	T1		**T2**		**T1**		**T2**		**T1**		**T2**		**T1**		**T2**	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
G1	.55	.50	.62	.48	.68	.47	.68	.47	2.12	1.46	1.97	1.35	2.56	1.84	2.25	1.62
G2	.60	.49	.62	.48	.61	.49	.64	.48	1.58	1.25	1.57	0.96	1.67	1.28	1.69	1.32
G3	.54	.50	.57	.49	.62	.49	.66	.47	1.50	1.01	1.91	1.44	2.19	1.67	2.28	1.69
G4	.57	.50	.52	.50	.64	.48	.64	.48	2.28	2.02	1.37	1.10	2.63	1.84	1.65	1.02
G5	.55	.50	.63	.48	.65	.48	.66	.48	1.51	1.22	1.57	1.41	1.55	1.04	1.62	1.46
G6	.62	.49	.68	.46	.63	.48	.63	.48	1.44	1.17	1.53	1.12	1.70	1.31	1.60	1.12
G7	.58	.49	.63	.48	.71	.46	.70	.46	1.71	1.45	1.47	1.02	2.28	1.71	1.89	1.49
G8	.54	.50	.60	.49	.68	.47	.64	.48	1.22	0.83	1.44	1.03	1.56	0.97	1.74	1.41
G9	.52	.50	.52	.49	.71	.45	.71	.46	1.60	1.13	1.99	1.43	1.53	1.31	1.95	1.56

For the DWR spectral distance scores, the GLMM revealed a significant main effect of *Testing Time,* $F(1,18050) = 23.480$, $p < 0.001$, and *Vowel,* $F(1,18050) = 11.358$, $p = 0.001$, and a significant *Testing Time* × *Group* interaction, $F(8,18050) = 7.996$, $p < 0.001$, and *Group* × *Vowel* interactions, $F(8,18050) = 3.018$, $p = 0.002$. Bonferroni-adjusted pairwise comparisons indicated that the *Testing Time* × *Group* interaction arose because three of the four groups trained with nonword stimuli (G1, G3 and G4) and only one of the four trained with word stimuli (G6, W + V + silence) produced both target vowels more accurately than the other groups.

For LD accuracy, the GLMM revealed a significant main effect of *Testing Time*, $F(1,6376) = 4.645$, $p = 0.031$, and a significant *Group* × *Vowel* interaction, $F(8,6376) = 2.652$, $p = 0.007$. None of the other fixed factors or interactions reached significance.

For the DSR spectral distance scores, no significant main effects were found, but the *Testing Time* × *Group,* $F(8,3708) = 10.488$, $p < 0.001$, and *Group* × *Vowel* interactions, $F(8,3708) = 3.956$, $p < 0.001$, turned out to be significant. Bonferroni-adjusted pairwise comparisons showed that only group G4 (NW + A + silence) produced the /æ/ significantly more accurately at post-test, as it was also the case in the DWR task.

Overall, the results show that the HVPT improved learners' discriminability of the L2 vowel contrast (ABX and DWR tasks), but little improvement was obtained in the lexical encoding of the contrast (DSR and LD tasks). Production gains were very modest, but groups trained with nonwords (G1, G2, G3, G4) gained significantly more than groups trained with words (G5, G6, G7, G8).

5.2 Attention Control and L2 Training Gains

Participants obtained a mean score of 94.60 (*SD* = 16.14, *Range* = 52–125) in the ASA task. In the ASW task, as expected, participants were significantly less accurate, $t(26206) = -7.326$, $p < 0.001$, and slower, $t(22771) = 30.759$, $p < 0.001$, on switch trials (Acc: $M = 0.88$, $SD = 0.326$; RT: $M = 976.44$ ms, $SD = 350.09$) than on repeat trials (Acc: $M = 0.91$, $SD = 0.290$; RT: $M = 840.53$ ms, $SD = 316.42$). Their attention switch-cost score ($M = 139.36$, $SD = 90.95$) was used in the correlation analyses.

Overall, correlational analyses failed to reveal an association between learners' gains in L2 vowel perception and production and the attention control measures, suggesting that gain sizes were unrelated to individual differences in attention control. Only a weak correlation, $r = 0.279$, $p = 0.004$, arose between ASA and DWR gains. Correlational analyses conducted separately by group yielded a similar picture. ASA was unrelated to any of the gain measures in all training groups. Nevertheless, ASW scores were strongly associated with some of the gain measures for some of the groups (Table 4).

Table 4 *Pearson-r* correlation coefficients between ASW and L2 perception and production gains (shaded cells indicate significance)

	/æ/							
	ABX		**LD**		**DWR**		**DSR**	
	r	*p*	*r*	*p*	*r*	*p*	*r*	*p*
G1	−.552	.050	−.237	.435	−.440	.133	−.390	.188
G2	−.096	.780	−.188	.580	−.418	.201	−.836	<.001
G3	.091	.768	−248	.413	.088	.775	−.242	.426
G4	.152	.604	.127	.665	−.420	.135	.016	.957
G5	.331	.270	−.279	.355	−.071	.817	−.423	.150
G6	−.713	.009	−.683	.014	.224	.484	−.014	.996
G7	.004	.988	.244	.053	−.538	.047	.346	.247
G8	−.099	.760	.193	.548	.266	.404	−.028	.931
	/ʌ/							
	ABX		**LD**		**DWR**		**DSR**	
	r	*p*	*r*	*p*	*r*	*p*	*r*	*p*
G1	−.185	.546	−.122	.691	−.577	.059	−.159	.603
G2	.027	.936	−.144	.673	−.345	.298	−.200	.555
G3	.595	.052	−.742	.004	.280	.354	−.374	.209
G4	−.223	.444	−.080	.787	−.582	.029	−.577	.039
G5	.140	.648	.227	.457	−.027	.929	.104	.734
G6	.028	.931	−.580	.050	−.049	.880	.252	.430
G7	.349	.221	−.034	.909	−.103	.725	−.033	.915
G8	.214	.505	−.036	.911	.002	.988	−.119	.713

- ASW explained gain differences in the production of /æ/ in the DSR task ($p <$ 0.001) for G2 (NW + V + silence).
- ASW was significantly correlated with gains in perceptual discrimination (ABX) ($p = 0.009$) and lexical encoding (LD) ($p = 0.014$) of /æ/ for G6 (W + V + silence).
- ASW explained a 55% of variance in the lexical encoding measure (LD) of /ʌ/ for G3 (NW + A + noise) and a 29% of variance in the production of words containing /æ/ for G7 (W + A + noise).
- Learners with stronger ASW skills in G4 (NW + A + silence) produced the L2 vowel /ʌ/ in the DWR and DSR significantly more accurately than those with poorer attention control (moderately strong correlations).

In sum, attention control (ASA and ASW) was not strongly related to gains in L2 vowel sensitivity and lexical encoding, but it helped in the conditions that required higher attentional demands (G2, G3, G6, G7).

Since as a whole attention control appeared to be unrelated to training gains, we explored whether it was related to individual differences in performance in the perception and production tasks at both testing times. Here we found that ASA was significantly related to ABX accuracy at T1 (/æ/: $r = 0.533$, $p < 0.001$; /ʌ/: $r = 0.508$, $p < 0.001$) and at T2 (/æ/: $r = 0.464$, $p < 0.001$; /ʌ/: $r = 0.473$, $p < 0.001$), explaining 21–28% of variance in participants' sensitivity to the target contrast, whereas ASW was only weakly related to ABX accuracy at T1 (/ʌ/: $r = -0.226$, $p = 0.022$). No significant associations were found between ASA or ASW and LD, DWR or DSR scores at T1 or T2. Therefore, ASA correlates strongly with ABX discrimination, which requires learners to perceptually discern between competing L2 vowel qualities by selecting one stimulus over another within every trial.

6 Discussion

Overall, HVPT was effective at improving trainees' discrimination of /æ/-/ʌ/ in perception and production (RQ1). Phonetically-oriented training through nonwords (unbiased by learners' lexical representations) led to larger gains in production than training through words, supporting previous findings (Ortega et al., 2021; Thomson & Derwing, 2016). However, trainees did not improve the lexical encoding of the contrast (RQ2). Longer HVPT combined with extended meaningful use of the L2 exploiting the target contrast in communicative tasks may be necessary for advanced learners to modify the lexical encoding of a phonological contrast.

Concerning the relationship between auditory attention control and L2 perception and production gains (RQ3), neither ASA nor ASW explained individual differences in training gains. In fact, we expected attention control to explain little variance in gains for groups that had obtained relatively small gains. Only ASW scores were found to be related to gains in L2 vowel learning, and only for some of the groups (G2, G3, G4, G6 and G7). It seems that learners' ability to switch between vowel quality and quantity explained learning gains especially for those who had been trained on either visual or background noise conditions. However, contrary to our expectations, ASW skills were unrelated to gains when learners were trained under the most demanding condition (visual monitoring + noise). Further research is needed to investigate this lack of relationship.

Concerning RQ4, ASA correlated strongly with learners' T1 and T2 scores in the ABX task, indicating that ASA enhanced learners' ability to discern between the target vowels, supporting previous findings (Mora & Mora-Plaza, 2019). However, neither ASA nor ASW were found to consistently interact with the training conditions in explaining gains, possibly due to training gains being relatively small within groups and testing not including any of the conditions implemented in the training. These findings suggest that further research should examine the role of attention control in learners' performance within training sessions from an individual differences perspective. Attention control may be more directly implicated in learners' actual training performance in perceptual discrimination and identification, as well as in the production tasks, during which the noise and visual monitoring conditions were present.

7 Pedagogical Implications

7.1 Implications for Phonetic Training

The present study demonstrates that HVPT helps learners better categorize vowels produced by different L2 speakers, and improves their L2 phonetic skills by helping them place the indexical information in the input (speakers' voice quality) in the perceptual background, thus enhancing the development of L2 phonetic categories during perceptual learning (Best, 2011). Moreover, HVPT may help learners develop pronunciation learning strategies in identifying new words from new speakers that can be transferred to production, thus contributing effectively to L2 pronunciation learning.

Pronunciation practice outside the laboratory could be provided through computer-assisted pronunciation training applications. These applications are designed to draw learners' attention to sounds and minimize attention to meaning, are interactive and entertaining, and involve immediate corrective feedback. For example, the *English Accent Coach* (Thomson, 2018), which was designed using a principled, research-based approach, showed to effectively improve pronunciation (Thomson, 2011). This website may improve speech comprehensibility and intelligibility without production practice. It also allows endless research possibilities as teachers and researchers could collaborate remotely, monitoring the effect of perceptual training and its impact on pronunciation.

7.2 Implications for Pronunciation Teaching

Cognitive attention control is likely to play an important role in the context of communicative language teaching. Meaning-oriented tasks where attention is directed to phonetic form have been shown to be effective in developing L2 speech perception and production skills (Gurzynski-Weiss et al., 2017).

Given that attention to phonetic features is necessary for pronunciation learning, teachers should ensure that students have as much exposure as possible to L2 speech that preserves phonological contrasts between L2 phonemes. One way of achieving this is to first provide explicit pronunciation practice through the use of nonwords (Mora & Levkina, 2017) and then progressively incorporate communicative tasks that require learners to use contrasting L2 sounds in real words (Tyler, 2019). Teachers could gradually change their focus-on-form tasks to real-world task-based pronunciation teaching tasks. This may be possible through the use of map tasks using words (Solon et al., 2017) or realistic problem-solving tasks that make the target phonological features essential for task completion and orient learners' attention to L2 phonological elements through the manipulation of task features (i.e., ±task complexity) (Mora-Plaza et al., 2018).

8 Conclusion

The present study has contributed to research on individual differences in L2 speech learning by exploring the role of auditory attention control in the phonetic training of L2 vowels. Based on prior research, it was hypothesized that training learners to exploit their attentional resources in phonetic form-focused pronunciation tasks to learn to perceive L2 phonological contrasts may prove a successful strategy to improve L2 pronunciation. Our study shows that Catalan-Spanish bilingual adult learners of English improved their ability to discriminate /æ/-/ʌ/ in perception and production tasks after receiving phonetic training, and that their production gains were larger when the training was through nonwords rather than through words. Yet, their lexical encoding of the contrast did not improve, and neither ASA nor ASW explained individual differences in training gains. Longer phonetic training with communicative tasks that draw attention to form may be necessary for advanced learners to modify the lexical encoding of a phonological contrast. For example, pair work involving minimal-pair based spot-the-difference tasks performed in noise might provide effective classroom training in auditory attentional skills that learners may find useful for L2 implicit perceptual learning through exposure to L2 oral input. Further research should empirically test the pedagogical value of manipulating auditory attentional demands to promote L2 pronunciation learning.

The present study is subject to several limitations. Sample sizes were small (11–14 per group). The visual monitoring and noise training conditions were implemented during production training only; they should have also been included during perception training. Finally, we tested production without visual monitoring or masking noise irrespective of training condition. In addition, as many of the target sources of individual differences are likely to be related to one another (e.g., auditory processing skills are likely to be related to cognitive attention control), it would be convenient to include as many potentially related variables in a single study as possible. This would allow researchers to statistically assess the joint and unique contribution of predictor variables while controlling for the confounding effects of mediating ones. Finally, further research is needed to investigate the role of attention control within each training session to observe whether attention plays a role during training.

References

Astheimer, L. B., Berkes, M., & Bialystok, E. (2016). Differential allocation of attention during speech perception in monolingual and bilingual listeners. *Language, Cognition and Neuroscience, 31*(2), 196–205. https://doi.org/10.1080/23273798.2015.1083114

Best, C. T. (2011). Devil or angel in the details? Perceiving phonetic variation as information about phonological structure. In J. Romero & M. Riera (Eds.), *The phonetics-phonology interface: Representations and methodologies* (pp. 3–32). John Benjamins. https://dx.doi.org/10.1075/cilt.335.01bes

Boersma, P., & Weenink, D. (2020). *Praat: Doing phonetics by computer* (Version 6.1.09) [Computer software].

Bohn, O. S., & Flege, J. E. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics, 11*(3), 303–328. https://doi.org/10.1017/S0142716400008912

Bradlow, A. R. (2008). Training non-native language sound patterns: Lessons from training Japanese adults on the English /r/-/l/ contrast. In J. G. Hansen Edwards & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 287–308). John Benjamins. https://dx.doi.org/10.1075/sibil.36.14bra

Carlet, A., & Cebrian, J. (2019). Assessing the effect of perceptual training on L2 vowel identification, generalization and long-term effects. In A. M. Nyvad, M. Hejná, A. Højen, A. B. Jespersen, & M. H. Sørensen (Eds.), *A sound approach to language matters: In honor of Ocke-Schwen Bohn* (pp. 91–119). Aarhus University. https://dx.doi.org/10.7146/aul.322.218

Cebrian, J. (2006). Experience and the use of duration in the categorization of L2 vowels. *Journal of Phonetics, 34*(3), 372–387. https://doi.org/10.1016/j.wocn.2005.08.003

Cebrian, J. (2019). Perceptual assimilation of British English vowels to Spanish monophthongs and diphthongs. *Journal of the Acoustical Society of America*, *145*(1), EL52–EL58. https://dx.doi.org/10.1121/1.5087645

Cebrian, J., Mora, J. C., & Aliaga-Garcia, C. (2011). Assessing crosslinguistic similarity by means of rated discrimination and perceptual assimilation tasks. In M. Wrembel, M. Kul, & K. Dziubalska-Kołaczyk (Eds.), *Achievements and perspectives in SLA of speech, New Sounds 2010* (Vol. 1, pp. 41–52). Peter Lang. ISBN: 9783631607220

Cooke, M., & Garcia-Lecumberri, M. (2018). Effects of exposure to noise during perceptual training of non-native language sounds. *The Journal of the Acoustical Society of America, 143*(5), 2602–2610. https://doi.org/10.1121/1.5035080

Darcy, I., Mora, J. C., & Daidone, D. (2014). Attention control and inhibition influence phonological development in a second language. *Proceedings of the 7th International Symposium on the Acquisition of Second Language Speech, New Sounds 2013: Concordia Working Papers in Applied Linguistics, 5*, 115–129. https://hdl.handle.net/2022/22863

Draine, S. (1999). *Inquisit* (Version 5.0.14.0) [Computer software]. Millisecond Software. https://www.millisecond.com/

Flege, J. E. (2009). Give input a chance! In T. Piske & M. Young-Scholten (Eds.), *Input matters in SLA* (pp. 175–190). Multilingual Matters.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, 35*(1), 116–124. https://doi.org/10.3758/BF03195503

Ghaffarvand Mokari, P., & Werner, S. (2019). On the role of cognitive abilities in second language vowel learning. *Language and Speech, 62*(2), 260–280. https://doi.org/10.1177/0023830918764517

Golestani, N., & Zatorre, R. J. (2009). Individual differences in the acquisition of a second language phonology. *Brain and Language, 109*(2–3), 55–67. https://doi.org/10.1016/j.bandl.2008.01.005

Gurzynski-Weiss, L., Long, A. Y., & Solon, M. (2017). TBLT and L2 pronunciation: Do the benefits of tasks extend beyond grammar and lexis? *Studies in Second Language Acquisition, 39*(2), 213–224. https://doi.org/10.1017/S0272263117000080

Hardison, D. M. (2018). Effects of contextual and visual cues on spoken language processing: Enhancing L2 perceptual salience through focused training. In S. M. Gass, P. Spinner, & J. Behney (Eds.), *Salience in second language acquisition* (pp. 201–220). Routledge.

Hazan, V., & Baker, R. (2011). Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions. *The Journal of the Acoustic Society of America, 130*(4), 2139–2152. https://doi.org/10.1121/1.3623753

Hazan, V., Sennema, A., Iba, M., & Faulkner, A. (2005). Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Communication, 47*(3), 360–378. https://doi.org/10.1016/j.specom.2005.04.007

Humes, L. E., Lee, J. H., & Coughlin, M. P. (2006). Auditory measures of selective and divided attention in young and older adults using single-talker competition. *The Journal of the Acoustical Society of America, 120*(5), 2926–2937. https://doi.org/10.1121/1.2354070

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English/r/-/l/to Japanese adults. *The Journal of the Acoustical Society of America, 118*(5), 3267–3278. https://doi.org/10.1121/1.2062307

Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., & Golestani, N. (2015). The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds. *The Journal of the Acoustical Society of America, 138*(2), 817–832. https://doi.org/10.1121/1.4926561

Kim, Y. H., & Hazan, V. (2010). Individual variability in the perceptual learning of L2 speech sounds and its cognitive correlates. In K. Dziubalska-Kołaczyk, M. Wrembel, & M. Kul (Eds.), *Proceedings of the 6th International Symposium on the Acquisition of Second Language Speech, New Sounds 2010* (pp. 251–256). Poznań, Poland. ISBN: 978-83-928167-9-9

Kormos, J. (2017). The effects of specific learning difficulties on processes of multilingual language development. *Annual Review of Applied Linguistics, 37*, 30–44. https://doi.org/10.1017/S026719051700006X

Lev-Ari, S., & Peperkamp, S. (2014). The influence of inhibitory skill on phonological representations in production and perception. *Journal of Phonetics, 47*, 36–46. https://doi.org/10.1016/j.wocn.2014.09.001

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes, 27*(7–8), 953–978. https://doi.org/10.1080/01690965.2012.705006

Meara, P. M., & Miralpeix, I. (2006). *Y_Lex: The Swansea advanced vocabulary levels test* (Version 2.05) [Computer Software]. Lognostics. https://www.lognostics.co.uk/tools/

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science, 21*(1), 8–14. https://doi.org/10.1177/0963721411429458

Mora, J. C., & Levkina, M. (2017). Task-based pronunciation teaching and research: Key issues and future directions. *Studies in Second Language Acquisition, 39*(2), 381–399. https://doi.org/10.1017/S0272263117000183

Mora, J. C., & Mora-Plaza, I. (2019). Contributions of cognitive attention control to L2 speech learning. In A. M. Nyvad, M. Hejná, A. Højen, A. B. Jespersen, & M. H. Sørensen (Eds.), *A sound approach to language matters: In honor of Ocke-Schwen Bohn* (pp. 477–499). Aarhus University. https://dx.doi.org/10.7146/aul.322.218

Mora-Plaza, I., Mora, J. C., & Gilabert, R. (2018). Learning L2 pronunciation through communicative tasks. In J. Levis (Ed.), *Proceedings of the 9th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, University of Utah, September, 2017 (pp. 174–184). Ames, IA: Iowa State University.

Moyer, A. (2014). What's age got to do with it? Accounting for individual factors in second language accent. *Studies in Second Language Learning and Teaching, 4*(3), 443–464. https://dx.doi.org/10.14746/ssllt.2014.4.3.4

Munro, M. J., & Bohn, O.-S. (2007). The study of second language speech: A brief overview. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience, second language learning: In honor of James Emil Flege* (pp. 3–11). John Benjamins. https://dx.doi.org/10.1075/lllt.17.06mun

Ortega, L., Iwashita, N., Norris, J. M., & Rabie, S. (2002, October 3–6). *An investigation of elicited imitation tasks in crosslinguistic SLA research* [Conference presentation]. Second Language Research Forum, Toronto, Canada.

Ortega, M., Mora-Plaza, I., & Mora, J. C. (2021). Differential effects of lexical and non-lexical high-variability phonetic training on the production of L2 vowels. In A. Kirkova-Naskova, A. Henderson, & J. Fouz-González (Eds.), *English pronunciation instruction: Research-based insights* (pp. 328–355). John Benjamins. https://dx.doi.org/10.1075/aals.19.14ort

Ou, J., Law, S. P., & Fung, R. (2015). Relationship between individual differences in speech processing and cognitive functions. *Psychonomic Bulletin & Review, 22*(6), 1725–1732. https://dx.doi.org/10.3758/s13423-015-0839-y

Rallo-Fabra, L., & Romero, J. (2012). Native Catalan learners' perception and production of English vowels. *Journal of Phonetics, 40*(3), 491–508. https://doi.org/10.1016/j.wocn.2012.01.001

Sadakata, M., & McQueen, J. M. (2013). High stimulus variability in nonnative speech learning supports formation of abstract categories: Evidence from Japanese geminates. *The Journal of the Acoustical Society of America, 134*(2), 1324–1335. https://doi.org/10.1121/1.4812767

Safronova, E., & Mora, J. C. (2013). Attention control in L2 phonological acquisition. In A. Llanes Baró, L. Astrid Ciro, L. Gallego Balsà, & R. M. Mateus Serra (Eds.), *Applied linguistics in the age of globalization* (pp. 384–390). Edicions de la Universitat de Lleida.

Saito, K., Kachlicka, M., Sun, H., & Tierney, A. (2020). Domain-general auditory processing as an anchor of post-pubertal L2 pronunciation learning: Behavioural and neurophysiological investigations of perceptual acuity, age, experience, development, and attainment. *Journal of Memory and Language, 115*, 104168. https://doi.org/10.1016/j.jml.2020.104168

Segalowitz, N., & Frenkiel-Fishman, S. (2005). Attention control and ability level in a complex cognitive skill: Attention shifting and second-language proficiency. *Memory & Cognition, 33*(4), 644–653. https://doi.org/10.3758/BF03195331

Shinohara, Y., & Iverson, P. (2018). High variability identification and discrimination training for Japanese speakers learning English/r/–/l. *Journal of Phonetics, 66*, 242–251. https://doi.org/10.1016/j.wocn.2017.11.002

Solon, M., Long, A. Y., & Gurzynski-Weiss, L. (2017). Task complexity, language-related episodes, and production of L2 Spanish vowels. *Studies in Second Language Acquisition, 39*(2), 347–380. https://doi.org/10.1017/S0272263116000425

Thomson, R. I. (2011). Computer assisted pronunciation Training: Targeting second language vowels: Perception improves pronunciation. *CALICO Journal, 28*(3), 744–765. https://dx.doi.org/10.11139/cj.28.3.744-765

Thomson, R. I. (2018). *English Accent Coach* (Version 2.3) [Computer software]. https://www.englishaccentcoach.com/

Thomson, R. I., & Derwing, T. M. (2016). Is phonemic training using nonsense or real words more effective? In J. Levis, H. Le., I. Lucic, E. Simpson, & S. Vo (Eds.), *Proceedings of the 7th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Dallas, TX, October 2015 (pp. 88–97). Ames, IA: Iowa State University.

Tyler, M. D. (2019). PAM-L2 and phonological category assimilation in the foreign language classroom. In A. M. Nyvad, M. Hejná, A. Højen, A. B. Jespersen, & M. H. Sørensen (Eds.), *A Sound approach to language matters: In honor of Ocke-Schwen Bohn* (pp. 607–630). Aarhus University. https://dx.doi.org/10.7146/aul.322.218

Wong, J. W. S. (2013). The effects of perceptual and/or productive training on the perception and production of English vowels /ɪ/ and /iː/ by Cantonese ESL learners. In F. Bimbot, C. Cerisara, C. Fougeron, G. Gravier, L. Lamel, P. Pellegrino, & P. Perrier (Eds.), *Proceedings of the 14th Annual Conference of the International Speech Communication Association, Interspeech 2013* (pp. 2113–2117). ISCA.

Ingrid Mora-Plaza is Ph.D. candidate and Lecturer in Phonetics and Phonology at the University of Barcelona, Spain. She is member of the Grup de Recerca en Adquisició de Llengües (GRAL) and L2 Speech Research groups and investigates the effectiveness of L2 phonetic training and inter-individual differences in proficiency, working memory, and auditory attention control. She is also interested in applying task-based principles to the teaching of L2 pronunciation.

Mireia Ortega is Lecturer in the Department of Modern Languages and Literatures at the University of Barcelona, Spain, where she obtained a Ph.D. in English Applied Linguistics. Her research interests include L2 pronunciation and teaching, phonetic training methods, crosslinguistic influence, and learners' individual differences. She is member of the Grup de Recerca en Adquisició de Llengües (GRAL) and L2 Speech Research groups.

Joan C. Mora is Associate Professor in the Department of Modern Languages and Literatures and English Studies at the University of Barcelona, Spain. His research has examined the acquisition of L2 phonology and the role of contextual and individual factors in the development of L2 speech and oral fluency.

The Effects of Intensive Phonetic Training on the Acquisition of English Stops

Ewelina Wojtkowiak

Abstract This chapter presents a longitudinal acoustic investigation of the effects of intensive phonetic training on the acquisition of English laryngeal contrasts (voiced vs. voiceless sounds) in consonant stops produced by Polish learners of English. To assess training effects, 55 L1 and 55 L2 realizations of initial /p, t, k/ and /b, d, ɡ/ mono- and di-syllabic words were collected from 10 Polish learners of English. The data revealed that the learners had acquired L2 aspirated voiceless stops before the phonetic training began and kept them separate from their L1 /p, t, k/ realisations throughout the study. The number of unvoiced, canonical "English-like" /b, d, ɡ/ productions skyrocketed from Time 1 to Times 2 and 3, suggesting that the phonetic training helped in successful pre-voicing suppression. However, the aspirates appeared to be the more "stable" category. The findings are interpreted in light of the Equivalence Classification Principle, which suggests that speakers should have fewer problems with the acquisition of a new category, which in this case is the aspirated stops. Voiced stops /b, d, ɡ/ are phonologically identical in both languages and thus are subject to more cross-linguistic interaction, which may explain why successful pre-voicing suppression rates are less permanent.

Keywords Phonetic training · Acoustic phonetics · Equivalence classification · Phonetic drift · Onset prominence

1 Introduction

Languages display striking differences with regards to voice onset time (VOT) (Lisker & Abramson, 1964). It is perhaps because of these differences and the fact that VOT is relatively easy to measure that laryngeal contrasts (fortis/voiceless vs. lenis/voiced consonants) are widely studied in L2 speech research. Based on VOT patterns, two-way systems are generally divided into two groups: *aspiration* languages and *true-voice* languages (Iverson & Salmons, 1995). The former group

E. Wojtkowiak (✉)
Adam Mickiewicz University, Poznań, Poland
e-mail: ewelina.wojtkowiak@amu.edu.pl

V. G. Sardegna and A. Jarosz (eds.), *Theoretical and Practical Developments in English Speech Assessment, Research, and Training*, Second Language Learning and Teaching,
https://doi.org/10.1007/978-3-030-98218-8_15

contrasts aspirated voiceless stops [pʰ, tʰ, kʰ] with lenis (i.e., unvoiced) stops [b̥, d̥, ɡ̊], as in the case of English or German. The latter group contrasts plain voiceless stops [p, t, k] with pre-voiced stops [b, d, g], as in the case of Polish and Spanish. (Note that henceforth the terms 'stop' and 'plosive' are used interchangeably.)

Figure 1 compares word-initial stops in English and Polish. VOT durations are highlighted in grey in the oscillograms. The VOT of the initial plosive consonant in the word *pub* produced by an English speaker as [pʰʌb̥] (top left) and by a Polish speaker as [pap] (bottom left) is 113 ms and 28 ms, respectively. Hence, the VOT in the aspirated plosive produced by the English speaker is longer than the VOT of the plain voiceless stop produced by the Polish speaker. In the voiced series, the English word *but* [b̥ˇʌt] (top right) displays a positive VOT value of 10 ms, while the Polish word *bat* 'a whip' [bat] (bottom right) displays a negative VOT value (i.e., voicing-lead) of −113 ms, followed by a burst, which indicates the release of the closure. Polish speakers of English who strive to attain native-like English pronunciation are required to do two things: first, they must acquire the long VOT associated with English aspirates and, second, learn to supress Polish-like pre-voicing in their L2 productions.

Keating (1984) notes that it is unclear how plain voiceless stops [p, t, k] differ from unvoiced lenis stops [b̥, d̥, ɡ̊]. Note, however, that the VOT of [p] in Polish *pub* is in

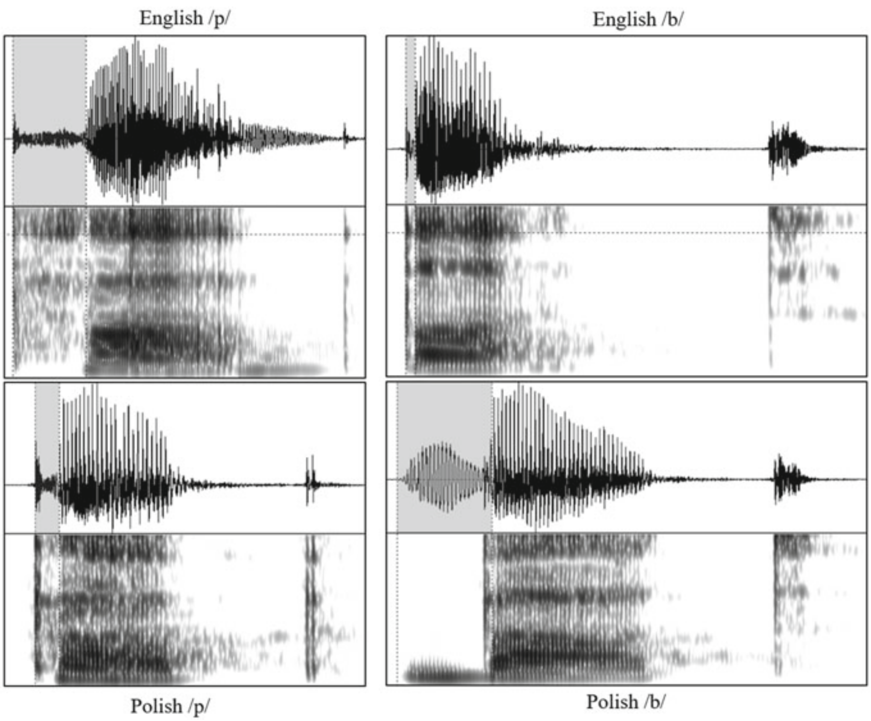

Fig. 1 VOT patterns (shaded area) in word-initial plosives in English and Polish

fact slightly longer than the positive VOT in English *but* even though both these words begin with bilabial plosives and are followed by a non-high vowel. Some studies, however, report that even though the difference in VOT duration between plain voiceless stops in Polish and unvoiced lenis stops in English is relatively robust, it still leads to a fair amount of confusion in discrimination studies (e.g., Kopczyński, 1977). Nonetheless, in her study of initial laryngeal voicing in Polish, Keating (1980) noted that there were overall remarkably few unvoiced realisations present in the productions of Polish monolingual speakers (see also Schwartz, 2020). On the other hand, pre-voicing has been attested for some accents of English (in particular, Southern American English; e.g., Hunnicut & Morris, 2016).

2 Literature Review

2.1 The Principle of Equivalence Classification

Many theories have been proposed to explain the differences between the realisations of laryngeal contrasts (e.g., CAH, Lado, 1957; PAM-L2, Best & Tyler, 2007; L2LP, Escudero, 2005). We chose to test the hypotheses formulated by the Speech Learning Model (SLM, Flege, 1995) because, unlike other models, SLM makes an explicit claim about the link between perception and production. In particular, Flege (1999) reports on a number of studies that looked at the correlation between perception and production and yielded a moderate outcome (i.e., $r = 0.50$). He suggests a few explanations as to why the correlation has not been found to be stronger. First, he notes that speakers might be able to correctly perceive an L2 sound and create a new phonetic category, but this ability might either not be extrapolated to production at all, or the implementation of the perceived contrast might take time. This outcome points to an advantage of longitudinal over cross-sectional studies: a speaker might be unable to accurately produce a sound at T1, but the result might be different at T2. In other words, "accurate perception does not entail accurate production whereas accurate production requires accurate perception" (Flege, 2016, p. 31). Second, SLM postulates the *Principle of Equivalence Classification*, which states that L1 affects L2 acquisition, but also L2 affects L1 production (phonetic drift and phonetic attrition in L1). In other words, SLM highlights the fact that both L1 and L2 exist in a common phonological space (Flege, 1995), and as such, the bidirectional influence is to be expected.

According to the *Principle of Equivalence Classification* (Flege, 1987), a key postulate put forward within SLM, when L1 and L2 phones are perceived as the same phonetic category, due to them being very similar, perceptual linkage—two sounds converging phonetically—often takes place because learners tend to tune out acoustic details. Consequently, L2 sounds that are phonetically similar to the learner's L1 native sounds will be subject to more L1 interference and, thus, might be expected to be produced with a certain degree of foreign accent. On the other hand, L2 phones that are different enough will not be categorised as belonging to an already existing L1 category, so a new phonetic category is likely to be formed instead. These phones are expected to display less L1 interference and native-like realisations thereof are presumed.

If we assume that Equivalence Classification is the source of L1 influence upon L2 (and vice versa) and, subsequently, want to provide a phonological account of our data, then we must first consider what different theories of laryngeal phonology see as equivalent. In other words, different phonological theories might make different predictions. In traditional feature theory (Chomsky & Halle, 1968), both aspirated voiceless and plain voiceless stops would be described by means of the feature [−voice], while both unvoiced and pre-voiced stops would bear the feature [+voice]. Therefore, from the point of view of this model, the interaction is predicted for both series of stops as both /p, t, k/ and /b, d, ɡ/ in either of the languages are phonologically the same. Laryngeal Realism (Honeybone, 2005), which is probably the dominating model designed to account for two-way laryngeal systems, provides a different prediction. Plain voiceless stops in true-voice languages, such as Polish, are unspecified for a feature ([Ø]), while pre-voiced stops are described in terms of a monovalent feature [voice]. In turn, aspirated voiceless stops in aspiration languages (e.g., English) are specified for feature [fortis], while short-lag /b, d, ɡ/ are left unspecified ([Ø]). Therefore, when you compare the phonological specifications of Polish and English, cross-linguistic interaction is predicted for neither series, as they are not phonologically equivalent.

The typology proposed by Laryngeal Realism is met with (relative) acceptance among linguists, mostly due to the transparency of the relationship between VOT and phonological encoding. It also goes in line with the typological frequency of plain voiceless stops; namely, the fact that plain voiceless stops are quite so common across languages may stem from the fact that they are phonologically unmarked (i.e., not specified for any features). However, these claims have been questioned. Vaux and Samuels (2005) criticise the assumption postulated by Laryngeal Realism's representations of plain voiceless stops and support their position with an overview of studies showing that children do not actually acquire plain voiceless stops first. Furthermore, they show that neutralisation does not always result in the unaspirated stop—in German the neutralised stop might be claimed to be aspirated. Aspiration also requires less articulatory precision as speakers do not have to fit their productions into a relatively short time-window. There are also studies pointing to the fact that [−voice] is an active phonological feature in some languages, contrary to what is suggested by Laryngeal Realism's unary representations (e.g., Bennet & Rose, 2017, on Moro; Wetzels & Mascaró, 2001, on Parisian French or Yorkshire English).

An alternative proposal is made by the Onset Prominence (OP) representational model (Schwartz, 2016), which is compatible with some of the assumptions of Laryngeal Realism, but also accounts for [−voice] being phonologically active. OP disposes of linear, segment-oriented representations; instead, in the framework manner of articulation, the structural and stop-vowel CV sequence is seen as a universal. This stems from the fact that CV constitutes the most common syllable

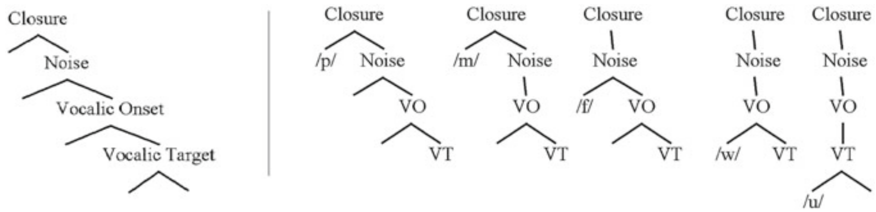

Fig. 2 OP representational hierarchy (on the left) and 'segments' extracted from the hierarchy (after Schwartz, 2016)

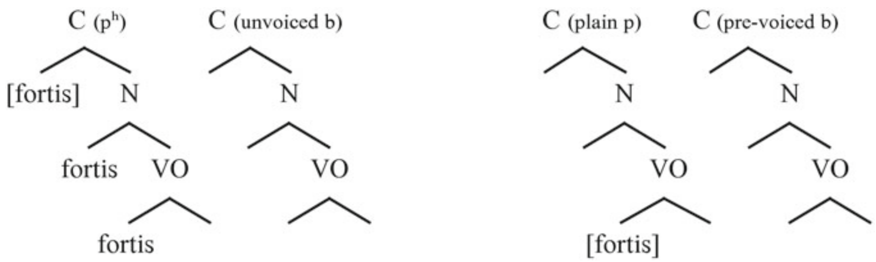

Fig. 3 OP representations of laryngeal contrasts in English (two left-most trees) and Polish (two right-most trees)

type across languages, and as such is taken as a phonological primitive (for a discussion, see Schwartz, 2016, p. 41). OP representations are shown in Fig. 2. On the left, we see the OP hierarchy built from a stop-vowel CV sequence, and on the right, we see individual 'segmental' structures extracted from the hierarchy. Each layer of the tree is associated with phonetic events taken out of the stop-vowel CV sequence. C is the closure phase in the production of a stop, the next level down is the N(oise) node, which corresponds to the release burst of the stop. The Noise node dominates the VO (Vocalic Onset) node, which is related to the initial portion of the vowel, often associated with formant transitions. VT (Vocalic Target) denotes the steady portion of the vowel target.

Initial laryngeal contrasts in OP, which are of importance for the purposes of the present chapter, depend on the presence or absence of the feature [fortis] in the specifications. Let us start with the representations of the voiceless series. Figure 3 depicts the difference between an aspirated voiceless plosive (first tree), such as those found in aspiration languages, versus a plain voiceless stop (third tree), such as those found in true-voice languages. Notice that in English (left) the feature [fortis] is assigned at the Closure level and then trickles down onto Noise and VO levels, which results in aspiration. In Polish, on the other hand [fortis] is assigned at the VO level, producing a plain voiceless stop.

Under OP, voicing is understood as the manifestation of a carrier signal, as envisioned in Modulation Theory (Traunmüller, 1994) and formalised in the OP framework (see Schwartz, 2017). The carrier signal is essentially the acoustic 'background,' which is a schwa-like vocoid and includes both the linguistic message as well as some extra-linguistic information, e.g., the speaker's attitude, age, sex. Therefore, the voiced series is left unspecified, as it does not modulate the carrier. Phonation is seen as part of the background, with [voice] not being a real phonological feature (see e.g., Cyran, 2014; Schwartz, 2020), and as a logical corollary, the voiced series is the unmarked series (cf. Schwartz, 2017 for a discussion). Figure 3 shows the voiced series (the second and the fourth trees). Notice that both unvoiced /b/ in English and pre-voiced /b/ in Polish are phonologically equivalent. Therefore, the prediction made by these representations is that Polish learners of English should display more cross-linguistic interaction in the voiced series than in the voiceless series. Aspiration, being phonologically different, results in a new category formation and, therefore, the aspirated and plain voiceless stops should be kept distinct.

2.2 The Acquisition of English Stops

The central issue of the present chapter is to determine the success rate with which Polish students acquire English stop consonants whilst undergoing pronunciation training. A question that arises at this point is whether different environments of L2 acquisition have any bearing on boosting the performance of the learners. In particular, it is intriguing to investigate whether formal instruction helps learners overcome the effects of equivalence classification and disentangle the L1 and L2 phones merged into one diphone, thus losing foreign accentedness. A number of studies has investigated to what degree phonetic training improves learners' L2 pronunciation (e.g., Champagne-Muzar et al., 1993; Couper, 2006; Derwing et al., 1997; Thomson, 2018). While the outcome of these studies was generally positive, the studies also portrayed methodological diversity of instruction. Production training in particular, especially when it is supplemented with regular language classes, has been shown to be beneficial for improving students' pronunciation (e.g., Mildner & Tomić, 2007).

When it comes to previous phonetic studies concerned with the acquisition of English laryngeal contrast by Polish learners, we observe a slight tendency: the voiced series remains relatively understudied. Waniek-Klimczak (2005) investigated the productions of /p, t, k/ by early and late Polish-English bilinguals and observed that both groups tended to produce values intermediate between Polish and English norms. Zając (2015) showed that Polish speakers converged phonetically with aspirated /p, t, k/, but had problems with pre-voicing suppression. Similar results were presented by Dzierla and Schwartz (2017), who showed high pre-voicing rates in the speech of very proficient speakers of English, with only ca. 40% of success rate in producing unvoiced, English-like items. These findings were replicated by Schwartz (2020) who compared two groups of speakers: 1st year students of English and Ph.D. students and professors from the same institution. He found that while, on the whole, both students and Ph.D. candidates/professors produced voiceless plosives with relatively long VOT (ca. 65 ms), pre-voicing was still prevalent in both groups' productions, with only slight decrease in duration for the latter group (ca.

−81.1 ms in students' realisations vs. −62.2 ms in Ph.D. candidates/professors'). Importantly, the successful pre-voicing suppression oscillated within chance level—students produced English lenis tokens with full, Polish-like, pre-voicing 58.6% of the time, while Ph.D. candidates/professors in 50.3% of the cases. Furthermore, Dzierla (2019) showed that perceptual training does not appear to affect the rates of pre-voicing suppression in the productions of Polish students of English. One surprising finding of his study was, however, that the voiceless category underwent improvement, even though it was not trained. In other words, while his training did not help participants lose pre-voicing in L2, it did help them acquire more native-like aspiration. In perception experiments, the relative weight of VOT as a cue at distinguishing English laryngeal contrasts has been questioned. Rojczyk (2011) found that Polish learners of English are not consistent at locating the boundary between aspirated and unaspirated stops in perception experiments. In turn, Schwartz and Arndt (2018) found that while Polish speakers displayed the highest accuracy when asked to distinguish between Polish pre-voiced /b, d, ɡ/ and voiceless /p, t, k/ (96.8%), they also did relatively well with regards to differentiating between Polish pre-voiced and unvoiced /b, d, ɡ/ tokens (75.3% of accurate responses). Aperliński (2012) further found that early learners of English attend to f_0 at vowel onset to a greater extent than to VOT as such when they are tasked with identifying laryngeal categories.

All in all, nonetheless, it appears that Polish learners of English have fewer problems with perception and subsequent production of aspirated English stops, which to them are more salient and different than what they find in their native language. In contrast, Polish learners of English have less control over pre-voicing.

3 The Study

The main objective of the present study was to explore cross linguistic interaction effects in the speech of Polish learners of English who are undergoing an intensive phonetic training during their first year of university education. Importantly, we set out to investigate how /p, t, k/ and /b, d, ɡ/ behave with regard to the assumptions of equivalence classification and whether there appeared to be any asymmetry in the behaviour, as found in previous phonetic research (e.g., Schwartz, 2020).

Our guiding research questions were the following:

1. Are there any equivalence classification effects in L2 productions of Polish learners of English?
2. Is the phonetic data compatible with the laryngeal typology proposed by the Onset Prominence framework?

It is hypothesised that equivalence classification will be observed for the voiced, but not voiceless series, which will yield support for OP representations. The prediction that stems from the OP framework is that our participants will keep the voiceless categories separate (i.e., a new category for aspirated English stops will be established and the VOT values for Polish and English plosives will be different) while the lenis/voiced series will be subject to much cross-linguistic interaction.

4 Methods

4.1 Participants

The participants in the experimental group were 10 first year students of English at Adam Mickiewicz University in Poznań, Poland. They were all female, aged 19–20 (median age: 19). Prior to the beginning of the study, all participants filled out a short background questionnaire. None reported any speech or hearing impairment.

Based on their questionnaire responses, they started learning English on average at the age of 6.5 years old (median age: 6). In high school, they attended on average 5.35 English lessons per week (median: 6, with one lesson lasting 45 min, which is a standard for Polish schools). Two participants mentioned that they took private tutoring lessons prior to their final high school exams (60 min per week in both cases) and the tutoring was conducted by native Polish teachers of English. Additionally, the students reported that neither their high-school teachers nor their tutors paid much attention to their English pronunciation. Moreover, no participant had ever had any classes with a native speaker of English, either in school or during private tutoring, and none had spent more than three consecutive weeks in an English-speaking country. Furthermore, and perhaps most importantly, none had undergone any form of a specialised phonetic training prior to the study.

4.2 The Nature of the Phonetic Training

As it is customary in the English programme at Adam Mickiewicz University, participants took two pronunciation-related annual courses: *Practical Phonetics* and *English Phonetics and Phonology*. The former was a practice-oriented course; the students attended two classes per week, 1.5 h each. The latter was more theory-driven and comprised one 45-min lecture per week, supplemented by one 90-min seminar per week, which dealt with the practical application of the theory presented during the lecture. Both courses were taught by six Polish native speakers with native-like fluency in English. All were trained phoneticians and phonologists and three of them actively conducted scientific research in the area of phonetics and phonology. Participants were enrolled in different groups but all attended the same courses with the same curricula. They all underwent the same phonetic training.

In relation to the English laryngeal contrast under study, most of the focus in both courses was decisively placed on teaching the differences between plain voiceless stops [p, t, k] and aspirated stops [p^h, t^h, k^h] (i.e., a voiceless stop at the beginning of a word or a stressed syllable which, in English, is produced with a little puff of air as the sound is released). Comparatively less focus was given to the lack of pre-voicing in English. There were numerous drilling exercises that taught aspiration; additionally, the students were required to provide the teachers with short recordings of texts from textbooks on pronunciation. One of the aspects that the teachers evaluated was whether or not the students employed aspiration. In *English Phonetics and Phonology*, the students were shown spectrograms that displayed the differences between both series of stops (voiced and voiceless) in Polish and English. When these contrasts were demonstrated, however, the students reported that they had no problem with perceiving English aspirated sounds, but they did not seem to hear any contrast between English pre-voiced stops /b, d, ɡ/, and unvoiced stops /p, t, k/. In addition, the students attended practical English classes with native and non-native speakers of English as well as some standard courses concerned with English and American history, literature, and culture. All classes were conducted exclusively in English.

One last remark that needs to be made with respect to the curriculum is that, in the second semester, the students were required to attend one foreign language course. The study participants took the following beginning-level foreign language courses: German (4 participants), French (4 participants), and Spanish (2 participants). Each of these courses consisted of 15 weekly meetings for 90 min. Yet, as the focus of instruction was mainly on reading and writing skills and on expanding the vocabulary and grammar of the foreign language, it is unlikely that any of these courses could have influenced the results of the current investigation.

4.3 Materials

The materials comprised two-word lists of initial /p, t, k/ and /b, d, ɡ/ mono- and di-syllabic words in English ($N = 55$) and in Polish ($N = 55$). They were counterbalanced for voicing (27 voiced and 28 voiceless stops in English, and 25 voiced and 30 voiceless stops in Polish) and place of articulation (21 bilabial /p, b/, 18 alveolar /t, d/ and 16 velar /k, g/ stops in English; and 17 bilabial, 19 alveolar and 19 velar stops in Polish). In both languages, the initial plosives were followed by a non-high vowel (/a, ɛ, ɨ, ɔ/ in Polish and /ɪ, e, æ, ɒ/ in English) in order to lower the risk of coarticulation processes influencing VOT duration (Keating, 1984; Klatt, 1975). Additionally, a number of fillers were included in the dataset. These fillers consisted of fricative- and cluster-initial mono- and di-syllabic items, licit in either of the languages. In total, the datasets (including fillers) comprised 132 words in English and 121 words in Polish.

4.4 Data Collection and Analysis

The students undergoing phonetic training were recorded six times in total (three testing times × two language sessions). The first recording (henceforth, T1) was made early in October, within the first two weeks of the academic year to ensure that no effects of phonetic training would be found yet. In other words, we wanted to record the productions of our students at the very onset of their university education to lower the risk of committing Type 1 error (finding a difference when there is not any)—that is, to trace the trajectory of the acquisition of the laryngeal contrast, we needed to ascertain the initial state of their English (and Polish) sounds. The second recording (henceforth, T2) was made in February, after one semester and during the winter exam session. The final recording (henceforth, T3) was held in early June, towards the end of the academic year, after completing approximately eight months of phonetic classes. Crucially, the recording sessions in Polish and English were kept separately (usually with a minimum of 24-h break) in order to avoid language mixing effects (Grosjean, 2004). They were conducted by a Polish native speaker, phonetically trained in English, and were held in the language about to be recorded (i.e., instructions given in Polish prior to recording the Polish words, and in English prior to recording the English words). The recordings were made in a sound-attuned booth at the university. They were recorded directly onto a laptop, using a condenser microphone and a USB interface. In total, 1650 target sounds were analysed in English and 1650 in Polish.

The items were elicited using PowerPoint slides, the order of which was pseudo-randomised and the same for each participant. They were subsequently analysed by hand in Praat (Boersma & Weenink, 2001). The acoustic parameter of interest was VOT. In the case of the voiceless series, we marked positive VOT, measured from the release of closure to the onset of voicing associated with the following vowel. In the case of the voiced series, we measured pre-voicing from the onset of voicing until the release burst (with the burst itself excluded). If pre-voicing was absent and short-lag positive was seen in the production of a voiced-initial item, we marked it as an English-like realisation. Therefore, we distinguished between two types of voiced stops: pre-voiced and unvoiced.

The statistical analysis was done in SPSS (IBM Corporation, 2019). A Generalised Linear Mixed Model was run with VOT as the dependent variable, separate for voiced and voiceless series. The interaction between *Session*Language*Type* was the main predictor, while *Speaker* and *Item* were included as random factors. Since the dataset was counterbalanced for place of articulation, this level was excluded from the analysis. An additional analysis was conducted for the type of VOT (coded as a binary variable: pre-voicing → yes or no) in the voiced series production; a Binary Logistic Regression was run with *Type* as the dependent variable, *Session* as the fixed factor, and *Speaker* and *Item* as random factors. A comparison group of Polish monolinguals was included in this analysis. Pairwise comparisons are reported in the next section.

5 Results

The Generalised Linear Mixed model revealed no effects of training across testing times for the English voiceless stops /p, t, k/ in the experimental group: T1 $M = 63.19$ ms ($SD = 27$), T2 $M = 60.69$ ms ($SD = 26$), and T3 $M = 61.90$ ms ($SD = 24$). *Mean* differences (*diff*) between testing times fell within 3 ms (T1 vs. T2 $diff = 2.23$ ms, $p = 0.118$; T2 vs. T3 $diff = 1.18$ ms, $p = 0.407$; and T1 vs. T3 $diff = 1.04$ ms, $p = 0.462$), which is well below the just-noticeable difference threshold. However, interestingly, as early as at T1, the students kept their English and Polish realisations of /p, t, k/ distinct: T1 $diff = 22.35$ ms ($SE = 1.610$, $t = 13.879$, $p = 0.000$); T2 $diff = 20.42$ ms ($SE = 1.610$, $t = 12.688$, $p = 0.000$); and T3 $diff = 23.74$ ($SE = 1.610$, $t = 14.790$, $p = 0.000$). This goes in line with the prediction of the students establishing a new category for English aspirates relatively easily. In contrast, the values of negative VOT duration of the voiced series /b, d, g/ do not indicate much difference across testing times between the English and Polish productions. Although the pre-voicing durations were slightly longer in L2 (English) productions, the differences between the two languages turned out not to be statistically significant: T1 $diff = 7.79$ ms ($SE = 5.125$, $t = 1.520$, $p = 129$), T2 $diff = 1.67$ ms ($SE = 5.799$, $t = 0.277$, $p = 0.782$), and T3 $diff = 9$ ms ($SE = 6.164$, $t = 1.463$, $p = 0.144$). Thus, the students seemed to show a lot of L1 interference in their English productions.

When it comes to pre-voicing realisations in English, much variation across and within speakers was found: T1 $M = -91.55$ ms ($SD = 26$), T2 $M = -84.69$ ms, ($SD = 26$), and T3 $M = -86.49$ ($SD = 29$). Phonetic training does not seem to have affected pre-voicing values in English production as the differences between the testing times were not significant (T1 vs. T2 $diff = 6.75$ ms ($SE = 5.548$, $t = 1.217$, $p = 0.224$); T1 vs. T3 $diff = 5.06$ ms ($SE = 5.992$, $t = 0.846$, $p = 0.398$); T2 vs. T3 $diff = 0.67$ ms ($SE = 6.412$, $t = 0.263$, $p = 792$).

The results for both series of stops are illustrated in Fig. 4.

Although a considerable amount of L1 interference was found, when it comes to type (i.e., the presence or absence of pre-voicing) of voiced series' realisation, we found some unvoiced, canonical English-like productions already in October ($N = 24$). The occurrence of lenis unvoiced productions got more numerous as training progressed, as there was observable progress between T1 ($N = 24$) and T2 ($N = 97$) ($SE = 0.064$, $t = 4.084$, $p < 0.001$) as well as between T1 and T3 ($N = 135$) ($SE = 0.074$, $t = 5.792$, $p < 0.001$) and T2 and T3 ($SE = 0.047$, $t = 3.583$, $p = 0.000$).

6 Discussion

The analysis of the English production of voiceless stops, /p, t, k/, produced by the experimental group indicated that there were no significant differences in VOT duration across the three testing times. However, a comparison between their Polish and English realisations suggested that the students acquired aspiration *before* they

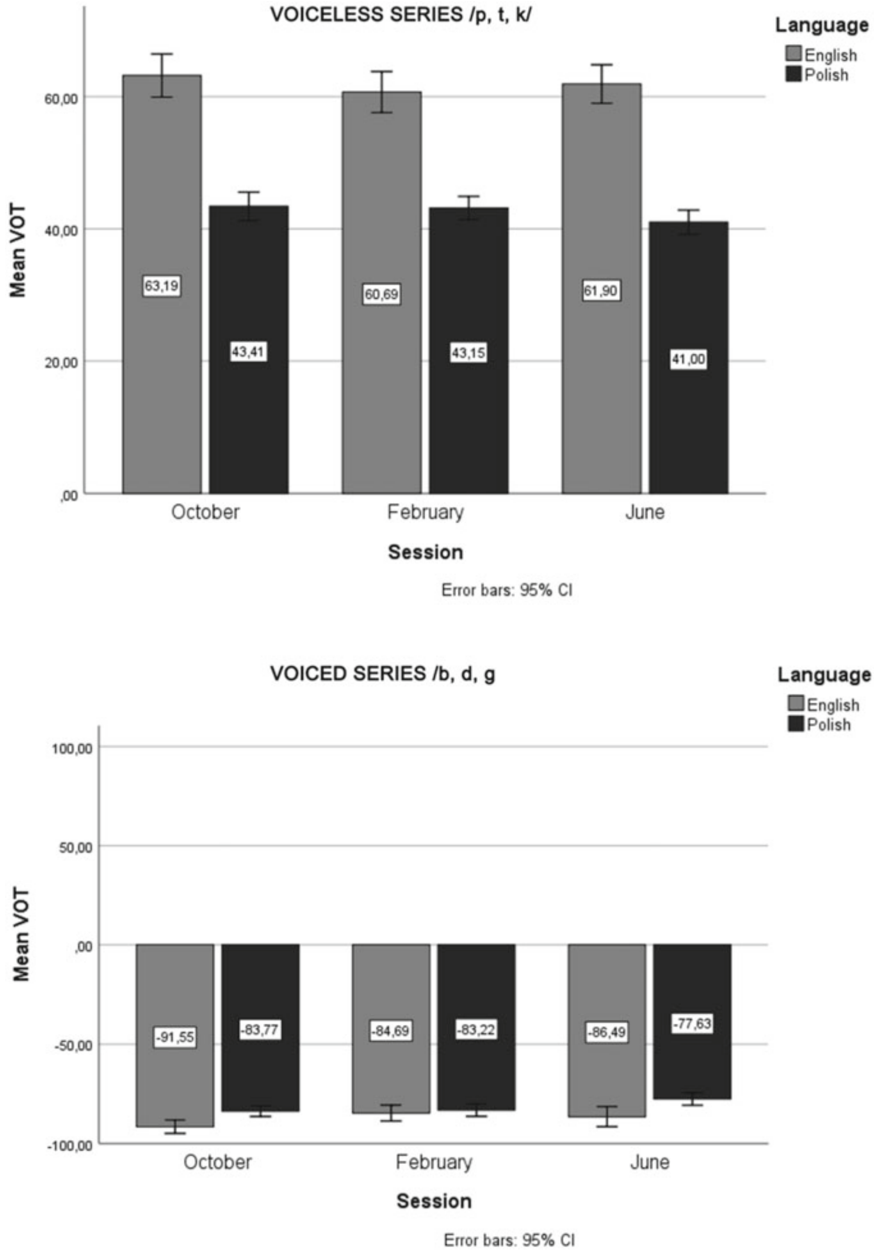

Fig. 4 Mean VOT values for voiceless (top) and voiced (bottom) stop realisations in English and Polish

began their phonetic training as the realisations of voiceless stops were significantly different in both languages at the three testing times. This difference might be explained by students' exposure to English prior to university education. Even though they reported not having had much interaction with native English speakers in conversation, they had been exposed to numerous English recordings at school as well as TV series and films, and it appears that when it comes to aspiration, this is enough for a new category to be formed. In other words, the students produced aspirated English sounds successfully despite not having that sound in their L1 repertoire. With respect to the voiced stops, /b, d, ɡ/, no difference in pre-voicing duration between the students' Polish and English productions was found. These findings lend support to the *Principle of Equivalence Classification* because more interaction was found in the category which was deemed identical by the speakers. Nonetheless, interestingly, as early as at T1, the students produced more English-like, unvoiced lenis realisations, therefore some interaction between English and Polish realisations of the voiced series was observed.

In general, both the L2 (English) and L1 (Polish) results suggest that the voiceless plosives, /p, t, k/, are more stable and less susceptible to the effects of cross-linguistic interaction than the voiced plosives, /b, d, ɡ/. Aspiration seems to be acquired without any problems and students are able to keep their VOT values of /p, t, k/ in L2 English distinct from L1 Polish. On the other hand, pre-voicing values differed greatly across (and within) our experimental group, in both the L1 and L2. While this longitudinal research shows that the learners made progress with respect to pre-voicing suppression, the acquisition of unvoiced, English-like items appears to take time and need explicit instruction, and has been found to be less permanent. After the training is over, pre-voicing sometimes makes a comeback whereas aspiration persists (cf. Schwartz, 2020). Furthermore, it may appear that our data contradict the findings of Dzierla (2019), whose perceptual training did not affect his students' /b, d, ɡ/ productions. However, it must be stressed that the training performed in that study was limited to very few sessions which did not last long and the pre-test and post-test data were gathered within approximately two weeks. For our study, the training lasted almost the full eight months. Therefore, it might be the case that it is more difficult—but not impossible—for Polish learners to produce voiced items with positive VOT values (see English /b/ vs. Polish /b/ in Fig. 1) and they simply need more time than a few weeks to show progress in the desired direction. This goes in line with Flege's (2016) argument that oftentimes production follows perception.

This type of asymmetry between voiceless and voiced stops with respect to cross-linguistic interaction has not been observed for Polish only. For instance, Schuhmann and Huffman (2015) observed that after receiving phonetic instruction, English students of Spanish showed a significant decrease in their L2 VOT productions of the voiceless series (of ca. 20 ms), which was a new category for them, while pre-voicing of /b, d, ɡ/ (found in the productions of some speakers) only approached significance. After the completion of their training, some pre-voicing was also found in their L1 English productions, while the VOT of /p, t, k/ stayed within monolingual norms.

In sum, it appears that while aspiration turned out to be found in the students' productions before the proper training began, the phonetic training they received might have boosted the performance of our students' voiced stops. After eight months, 50.75% of the voiced-initial tokens were produced with English-like values. However, it appears that the suppression of pre-voicing, while doable, proved to be rather difficult for the learners, as 49.25% of the productions at T3 were still realised with Polish-like negative VOT values. These results replicate previous findings on pre-voicing suppression (e.g., Dzierla & Schwartz, 2017; Schwartz, 2020).

7 Pedagogical and Research Implications

The findings of the present study point to an asymmetry with respect to the difficulty with which Polish learners acquire English laryngeal contrasts. It is possible that the asymmetry stems from the nature of the phonetic training. Indeed, aspiration appears to be the feature of more importance than the presence or absence of pre-voicing. The reason for that might be the fact that, for the purposes of communication, aspiration is much more crucial, while pre-voicing is by and large less perceivable. Nonetheless, the question whether or not more focus should be put on teaching students how to suppress pre-voicing is secondary; it does not have bearing on communication after all. From the perspective of OP, it is a phonetic detail without phonological significance. Pre-voicing is important as its variability and unstable nature provides empirical support to the phonological representations postulated by the OP framework.

Further research implications include the necessity of investigating *both* series of stops; focusing on the voiceless series only, and studying, for instance, the success with which aspiration is acquired (due to its saliency) might prove to be myopic in that it does not allow us to get the full picture of the trajectory of L2 acquisition.

8 Conclusion

This chapter attempted to investigate the success with which Polish learners of English acquire L2 laryngeal contrasts. In light of the predictions made by phonological theories with respect to equivalence classification effects in the acquisition of L2 stops, the framework that best accounts for the phonetic data discussed herein is OP. Recall that OP assumed that the voiced series is identical in both Polish and English, whereas the representation of aspirated vs. plain voiceless stops is different. Therefore, it is the voiced series that should present more cross-linguistic interaction. This is exactly what we found in our data—Polish and English /p, t, k/ were not subject to much L1 interference or L2-induced phonetic drift. Being a new category, aspiration also appeared to be acquired relatively early and easily. On the other hand, since /b, d, ɡ/ are phonologically equivalent, more interaction between English unvoiced and Polish pre-voiced realisations was expected. Indeed, we observed substantial L1 interference in L2 productions—after eight months of training, only 50.75% of the tokens were realised with English-like short-lag VOT.

Pre-voicing, then, appears to be merely a manifestation of the carrier signal, and as such is simply a phonetic detail which has no bearing on phonology (Schwartz, 2020). Other acoustic cues are much more important to maintaining a laryngeal contrast in Polish (in particular F1 at the onset of the vowel following the initial stop; Schwartz & Wojtkowiak, 2018). To the best of our knowledge, OP is the only model which predicts and accounts for the asymmetries in the behaviour of both series of stops.

Possible limitations of the study include a lack of an English control group, with which the L2 productions could be compared. Future investigation will attempt to rectify that. We also want to assess whether aspiration is present in longer utterances, not merely in citation forms. Nonetheless, the longitudinal data obtained for the purposes of this chapter, especially as both voiced and voiceless consonants are encompassed, yield empirical support to both the *Equivalence Classification Principle* and the phonological representations postulated by OP.

Acknowledgements Parts of this research have been supported by the Polish National Science Centre (Narodowe Centrum Nauki), project number UMO-2016/21/B/HS2/00610. Special thanks go to Katarzyna Dziubalska-Kołaczyk and Geoffrey Schwartz for their helpful advice.

References

Aperliński, G. (2012, December 6–8). *Is VOT enough?* [Conference presentation]. International Conference on Native and Non-Native Accents of English (Accents 2012), Łódź, Poland.

Bennet, W., & Rose, S. (2017). Moro voicelessness dissimilation and binary [voice]. *Phonology, 34*(3), 473–505. https://doi.org/10.1017/S0952675717000252

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O. S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13–34). John Benjamins. https://benjamins.com/catalog/lllt.17.07bes

Boersma, P., & Weenink, D. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5*(9/10), 341–345.

Champagne-Muzar, C., Schneiderman, E., & Bourdages, J. S. (1993). Second language accent: The role of the pedagogical environment. *IRAL: International Review of Applied Linguistics in Language Teaching, 31*(2), 143–160.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. Harper & Row.

Couper, G. (2006). The short and long-term effects of pronunciation instruction. *Prospect, 21*(1), 46–66. Retrieved January 29, 2022, from http://www.ameprc.mq.edu.au/__data/assets/pdf_file/0007/229831/21_1_3_Couper.pdf

Cyran, E. (2014). *Between phonology and phonetics: Polish voicing*. De Gruyter Mouton. https://doi.org/10.1515/9781614515135

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

Dzierla, J. (2019, June 25–26). *The effects of auditory training on the perception and production of English word-initial laryngeal contrasts by Poles* [Conference presentation]. Approaches to Phonology and Phonetics (APAP2019), Lublin, Poland.

Dzierla, J., & Schwartz, G. (2017, June 19–21). *Pre-voicing suppression in the speech of Polish learners of English* [Conference presentation]. Approaches to Phonology and Phonetics (APAP2017), Lublin, Poland.

Escudero, P. (2005). *Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization* [Doctoral dissertation, Utrecht University]. LOT Dissertation Series 113.

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, 14*(1), 47–65. https://doi.org/10.1016/S0095-4470(19)30537-6

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. 229–273). York Press.

Flege, J. E. (1999). The relation between L2 production and perception. *Proceedings of the 14th International Congress of Phonetic Sciences* (pp. 1273–1276). International Phonetic Association. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS1999/papers/p14_1273.pdf

Flege, J. E. (2016, June 10–12). *The role of phonetic category formation in second language speech acquisition* [Conference presentation]. The 8th International Conference on Second Language Speech (New Sounds 2016), Aarhus, Denmark. Retrieved January 29, 2022, from http://www.jimflege.com/files/Flege_Aarhus_talk_20162.pdf

Grosjean, F. (2004). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition, 1*(2), 131–149. https://doi.org/10.1017/S136672899800025X

Honeybone, P. (2005). Diachronic evidence in segmental phonology: The case of obstruent laryngeal specifications. In M. van Oostendorp & J. van der Weiher (Eds.), *The internal organization of phonological segments* (Vol. 77, pp. 319–354). De Gruyter Mouton. https://doi.org/10.1515/9783110890402

Hunnicut, L., & Morris, P. A. (2016). Prevoicing and aspiration in Southern American English. *Proceedings of the 39th Annual Penn Linguistics Conference, 22*(1), 24, 215–224. https://repository.upenn.edu/pwpl/vol22/iss1/24

IBM Corporation. (2019). *IBM SPSS statistics for Windows* (Version 26.0) [Computer software]. IBM Corp.

Iverson, G. K., & Salmons, J. C. (1995). Aspiration and laryngeal representation in Germanic. *Phonology, 12*(3), 369–396. https://doi.org/10.1017/S0952675700002566

Keating, P. A. (1980). *A phonetic study of a voicing contrast in Polish* (Publication No. 8111121) [Doctoral dissertation, Brown University]. ProQuest Dissertations Publishing.

Keating, P. A. (1984). Phonetic and phonological representation of stop consonant voicing. *Language, 60*(2), 286–319. https://doi.org/10.2307/413642

Klatt, D. H. (1975). Voice-onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research, 18*(4), 687–703. https://doi.org/10.1044/jshr.1804.686

Kopczyński, A. (1977). *Polish and American English consonant phonemes: A contrastive study*. Państwowe Wydawnictwo Naukowe.

Lado, R. (1957). *Linguistics across cultures: Applied linguistics and language teachers*. University of Michigan Press.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word, 20*(3), 384–422. https://doi.org/10.1080/00437956.1964.11659830

Mildner, V., & Tomić, D. (2007). Effects of phonetic speech training on the pronunciation of vowels in a foreign language. *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 1665–1668). Saarland University, Saarbrücken. Retrieved January 25, 2022, from http://www.icphs2007.de/

Pavlenko, A., & Jarvis, S. (2002). Bidirectional transfer. *Applied Linguistics, 23*(2), 190–214. https://doi.org/10.1093/applin/23.2.190

Rojczyk, A. (2011). Perception of time continuum by Polish learners. In J. Arabski & A. Wojtaszek (Eds.), *The acquisition of L2 phonology* (pp. 37–58). Multilingual Matters.

Schuhmann, K. S., & Huffman, M. K. (2015). L1 drift and L2 category formation in second language learning. *Proceedings of the 18th International Congress of Phonetic Sciences*. International Phonetic Association. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0850.pdf

Schwartz, G. (2016). On the evolution of prosodic boundaries: Parameter settings for Polish and English. *Lingua, 171*, 37–73. https://doi.org/10.1016/j.lingua.2015.11.005

Schwartz, G. (2017). Formalizing modulation and the emergence of phonological heads. *Glossa: A Journal of General Linguistics, 2*(1), 81–101. https://doi.org/10.5334/gjgl.465

Schwartz, G. (2020). *Asymmetrical cross-language phonetic interaction: Phonological implications*. Advance online publication. https://doi.org/10.1075/lab.19092.sch

Schwartz, G., & Arndt, D. (2018). Laryngeal realism vs. modulation theory: Evidence from VOT discrimination in Polish. *Language Sciences, 69*, 98–112. https://doi.org/10.1016/j.langsci.2018.07.001

Schwartz, G., & Wojtkowiak, E. (2018, October 11–13). *Acoustic evidence and representations for fortisness in Polish stops* [Conference presentation]. SinFonIJa 11, Kraków.

Thomson, R. I. (2018). High variability [pronunciation] training (HVPT): A proven technique about which every language teacher and learner should know. *Journal of Second Language Pronunciation, 4*(2), 208–231. https://doi.org/10.1075/JSLP.17038.THO

Traunmüller, H. (1994). Conventional, biological, and environmental factors in speech communication: A modulation theory. *Phonetica, 51*, 170–183. https://doi.org/10.1159/000261968

Vaux, B., & Samuels, B. (2005). Laryngeal markedness and aspiration. *Phonology, 22*(3), 395–436. https://doi.org/10.1017/S0952675705000667

Waniek-Klimczak, E. (2005). *Temporal parameters in second language speech: An applied linguistic phonetics approach*. Wydawnictwo Uniwersytetu Łódzkiego.

Wetzels, W. L., & Mascaró, J. (2001). The typology of voicing and devoicing. *Language, 77*(2), 207–244. https://doi.org/10.1353/lan.2001.0123

Zając, M. (2015). *Phonetic convergence in the speech of Polish learners of English* [Unpublished doctoral dissertation]. The University of Łódź.

Ewelina Wojtkowiak is a Ph.D. student in the Department of Contemporary English Language at the Faculty of English, Adam Mickiewicz University, Poland. Her research interests include phonological representation and the phonetics-phonology interface. She is currently working on cross-linguistic interaction effects in the speech of Polish learners of English, with special focus on phonetic drift. She specialises in acoustic phonetics.

The Acquisition of Phonology in Multilinguals

Natural Growth Theory of Acquisition (NGTA): Evidence from (Mor)Phonotactics

Katarzyna Dziubalska-Kołaczyk and Magdalena Wrembel

Abstract This chapter proposes a new explanatory framework for the acquisition of phonology and provides support for it with evidence from phonotactics and morphonotactics. Several models have been proposed to account for bilingual acquisition of speech; however, an optimal explanatory framework that would account for the complexity of the acquisition process is still lacking. We offer a new theory, called the Natural Growth Theory of Acquisition (NGTA), which is informed by natural phonology and complexity theory. NGTA assumes a gradual dynamic emergence of Ln phonology, shaped by input from the first language (L1) and other languages (Ls), and influenced by typology, universals, and context. It considers the universal, typological, and language-specific aspects of the growth. General assumptions of NGTA as well as induction from speech data let us formulate a catalogue of hypotheses concerning the acquisition of clusters. The hypotheses are corroborated by findings from four studies and explained by means of the linguistic and extralinguistic variables considered by NGTA.

Keywords Bi(multi)lingual phonological acquisition · Theory of acquisition · Phonotactics · Morphonotactics

1 Introduction

Different models have been proposed to account for bilingual acquisition of speech yet only a few assume a multilingual perspective. This chapter provides a critical assessment of relevant theoretical approaches to foreign language (FL) acquisition

K. Dziubalska-Kołaczyk · M. Wrembel (✉)
Adam Mickiewicz University, Poznań, Poland
e-mail: magdala@amu.edu.pl

K. Dziubalska-Kołaczyk
e-mail: dkasia@amu.edu.pl

V. G. Sardegna and A. Jarosz (eds.), *Theoretical and Practical Developments in English Speech Assessment, Research, and Training*, Second Language Learning and Teaching,
https://doi.org/10.1007/978-3-030-98218-8_16

and offers a new theory, called the Natural Growth Theory of Acquisition (NGTA), as a more optimal explanatory framework that can account for the complexity of the acquisition process in a multilingual mind. NGTA is informed by natural phonology and enhanced by complexity theory. It is holistic in the sense that it incorporates each and every aspect of the acquisition process. It assumes a gradual dynamic emergence of Ln phonology, shaped by input from the first language (L1) and other languages (Ls) and influenced by universal preferences (understood as preferability generalizations, see Sect. 3.2), typology, and context. NGTA is conceived as a general theory of language acquisition, which allows to model the acquisition of phonology as well as morphology and other language domains; yet in the present chapter we focus on speech and interpret the theory with generated data.

2 Overview of SLA Speech Models

Current models of second language (L2) phonology are mainly concerned with the relationship between the L1 and L2 of the speaker, the role of language universals, and the influence of non-linguistic factors on the rate, process, and outcome of phonological acquisition (Gut et al., 2015). Conversely, research into third language (L3) phonology is a very young discipline that has its roots in the late twentieth and early twenty-first centuries. The major difference is that L3/Ln learners have already acquired their first foreign language (L2). Thus, L3/Ln learners can resort to previous linguistic knowledge and language-learning experiences, and have a broader phonetic repertoire, enhanced perceptual sensitivity, and metalinguistic awareness, which may facilitate their learning of a subsequent phonological system (Gut, 2010; Wrembel, 2012).

Several approaches modelling the acquisition of speech have become prominent among second language acquisition (SLA) researchers, including the models proposed by Flege (1995), Best (1995), Pisoni (1996), and Kuhl and Iverson (1995). First, in Flege's (1987, 1995) Speech Learning Model (SLM), L1 and L2 position-sensitive allophones are related along a continuum of inter-lingual phonetic similarity defined in acoustic–phonetic terms, such as F1/F2 for vowels or VOT for consonants. Beginners perceptually assimilate most L2 categories to native ones. If the L2 segment is sufficiently dissimilar, a new L2 perceptual category is established over time. New category formation may be blocked by equivalence classification for less dissimilar sounds, so a single perceptual category subsumes both L1 and L2 segments, leading to persistent accented production in the L2 or even to shifts in L1 production. Equivalence classification is defined by Flege (1987) as "a basic cognitive mechanism which permits humans to perceive constant categories in the face of the inherent sensory variability found in the many physical exemplars which may instantiate a category" (p. 49). He hypothesizes that "equivalence classification prevents adult L2 learners from establishing a phonetic category for similar but not new L2 phones" (p. 50). For instance, he maintains that advanced native English speakers of French should produce the new French vowel /y/ accurately, but not the similar

French vowel /u/. Importantly, he also claims that the effect of interference between the L1 and L2 is bidirectional, since "L2 learners 'merge' the phonetic properties of similar L1 and L2 phones within a single category" (p. 51). Second, the Perceptual Assimilation Model (PAM, Best, 1995; PAM-L2, Best & Tyler, 2007) is concerned with initial perceptual difficulties. In this model, non-native phonetic segments are perceptually assimilated to native phonetic categories according to their articulatory similarity to the native gestural constellations. Perceptual difficulty in differentiating non-native contrasts is predictable from these assimilation patterns (cf. "goodness of fit" to a given native category). Third, in the Exemplar-Based Model (Pisoni, 1996), native phonetic categories are represented as clusters of exemplars that share certain critical acoustic parameters. Categorization involves matching an incoming signal to previously stored exemplars. In this model, L2 perceptual training should be conducive to the formation of (new) equivalence clusters. Finally, in the Native Language Magnet (NLM) Model (Kuhl & Iverson, 1995), native-language phonetic categories are organized around prototypes, established within the first year of life, which distort the phonetic perceptual space. L2 perceptual learning would require the reorganization of the phonetic perceptual space around newly established prototypes (for more comprehensive overviews of SLA speech models, see Leather & James, 1991; Strange, 1999).

Overall, the four previously discussed models of L2 speech acquisition focus on phonetic rather than phonological categories. It is not phonemes or features, but context-dependent phonetic segments that form the level of analysis in SLM and PAM, with clusters of exemplars and prototypes in NLM and Pisoni's Exemplar-Based Model. All four models rely on the notion of phonetic similarity (see Strange, 1999, for further discussion). In general, they make no specific claims with reference to how phonology works in the L2 learner's mind—that is, they do not describe processes or representations. The absence of such descriptions may be the reason why these four models do not readily fit a multilingual context, which requires a more global and holistic image of how phonology works in the multilingual speaker's mind.

However, SLM, PAM, NLM and Pisoni's Exemplar-Based Model may still be considered as phonological models if we refer to the interpretation of what is phonetic and what is phonological recently proposed by Chang (2019). According to Chang (2019), there is a distinction between phonological (phonotactic) and phonetic (cue-centric) transfer. He understands phonological transfer as the influence of L1 phonological constraints on L2 perception, while phonetic transfer is seen as the influence of relative functional load (RFL) of phonetic cues, such as VC transitions or vowel duration, based on perceptual attention. Consequently, phonological and phonetic hypotheses lead to different predictions. One of the outstanding questions recognised by scholars is how the L1 and L2 phonologies interact in influencing L3 perception and production. As a direction for future research, Chang (2019) proposes to examine systematically the interaction of L2 transfer with L1 transfer and with universal processes in L3 perception.

It has to be noted, however, that the L3 acquisition models proposed so far stem from and concentrate on morphosyntax (see Wrembel, 2015, for a detailed discussion), but no model, to the best of our knowledge, accounts for the acquisition of

L3 phonology. Therefore, we intend to fill in this gap and offer a wider perspective accounting for multilingual acquisition of speech.

3 Natural Growth Theory of Acquisition: A Proposal

This section aims to present a new explanatory framework of multilingual acquisition of speech and support it with data. We claim that there is a need for a holistic theory of language, which provides a big picture rather than a collection of observable or elicited details. We need a theory that would explain acquisition in all relevant aspects (i.e., L1, L2, L3, cross-linguistic influence, language attrition, and death) and allow for modelling the acquisition of speech. Further, this theory should be interdisciplinary and open to transdisciplinarity. With these goals in mind, we hereby propose the Natural Growth Theory of Acquisition (NGTA). For a preliminary version of this model, see Dziubalska-Kołaczyk and Wrembel 2017; Wrembel and Dziubalska-Kołaczyk 2016. In the next subsections, we first discuss the epistemological background compatible with NGTA and then present our epistemological stand. Next, we formulate NGTA's general assumptions.

3.1 Epistemological Background

The new explanatory theory—that is, NGTA—is compatible with natural phonology (Donegan & Stampe, 2009; Dressler, 1984, 1996; Dziubalska-Kołaczyk, 2002, 2009, 2012) and is enhanced by complexity theory (Burkette & Kretzschmar, 2018; Bybee, 2001, 2007, 2010; Kretzschmar, 2015). Natural phonology and complexity theory adopt different epistemological approaches to explaining linguistic phenomena. As part of natural linguistics, natural phonology derives preferences from general, higher principles such as the *Principle of Figure-and-Ground* or the *Principle of Economy*. In contrast, complex systems match preferences with the highest frequency of use and, thus, experience with forms. NGTA builds upon these two theoretical frameworks. Our goal is to modify and extend the explanatory potential offered by classic natural phonology and modern natural linguistics with epistemological support from complex systems and with inductive support from speech data, which we will elaborate on in the subsequent sections.

3.1.1 Insights from Natural Phonology

Natural phonology claims that phonological processes reflect real constraints on speaker abilities (Donegan & Stampe, 2009; Stampe, 1979), affecting both perception and production. It also maintains that a learner copes with the difficulties by modifying universal processes (which are either innate or developmentally available) to fit L1 phonology, later Ln phonologies. Those modifications are the limitations that a speaker brings to the language. System-internal and external criteria co-determine the difficulty of perception, production, and eventually acquisition of sounds.

Natural phonology itself has roots in the analysis of acquisition of phonetic representations (Dziubalska-Kołaczyk, 1990; Stampe, 1969). It provides an often missed link between phonology and phonetics by claiming that phonological systems are phonetically motivated (hence, natural). It takes into account cognitive and extralinguistic factors. As part of natural linguistics, natural phonology complies with the self-organization and dynamic emergence of structures, operates with preferences and graduality, takes into account frequency effects, and invokes semiotic principles, such as the *Principle of Figure-and-Ground*, the *Principle of Contrast*—both principles having psychological origins (e.g., Rubin, 1915)—and the *Principle of Cognitive Economy,* and the *Principle of Least Effort* (e.g., Zipf, 1949).

In 1990, Dziubalska-Kołaczyk proposed a model of L2 language phonology grounded in natural phonology (see Fig. 1).

In Dziubalska-Kołaczyk (1990)'s model, the learner acquires an L2 phonology by means of learning (as aided by instruction), which is a mechanism quite different from the automatic and uncontrolled acquisition that takes place in natural settings. L2 surface realisations are initially filtered through L1-specific sound intentions. In the

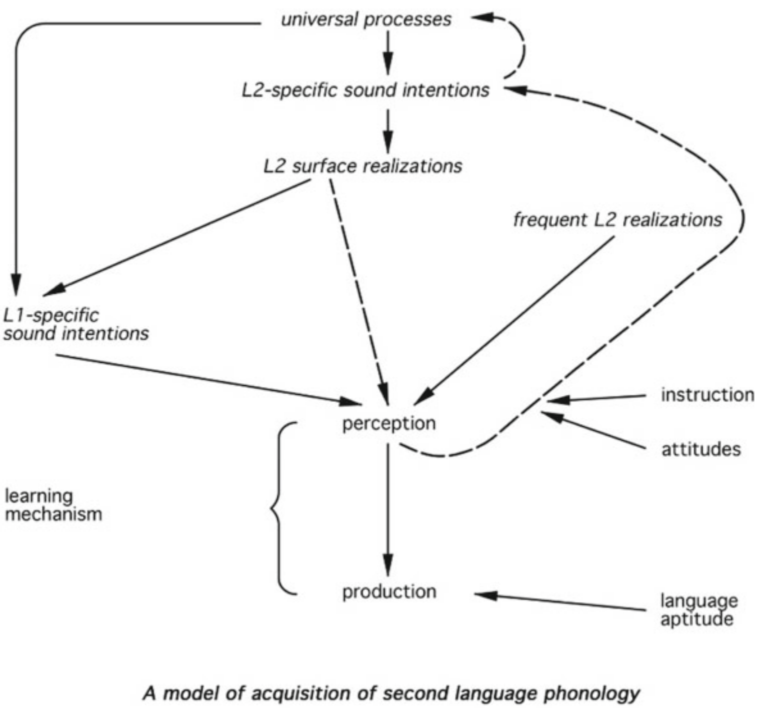

Fig. 1 A model of acquisition of L2 phonology

process of learning, the learner manages to uncover the L2-specific sound intentions that make him/her closer to the L2 target. The process is aided by exposure to frequent L2 realisations. Both L1 and L2-specific sound intentions stem from and feed back to universal processes shared by all speakers. Learning may ultimately result in a total *unsuppression* and correct limitation of natural phonological processes of a pre-linguistic stage (i.e., those natural processes selected to operate in the language learned). Success, however, may be expected only if all the conditioning factors of acquisition are favourable.

As the model describes, in search for an L2-specific sound intention, the learner observes and tries to imitate the foreign output. A classroom setting learner is trained to observe and imitate through instruction (there is no such formal training for a natural setting learner). Also, learners of the two settings are differentiated by social-psychological factors in the background of acquisition. Learner attitudes and, consequently, motivation as well as language aptitude may differ quite substantially in the two settings, which may, in turn, lead to an essential difference in learner achievement/L2 production. Neither is the *learning* procedure the same as that of the child: the latter cannot, in principle, be driven in the acquisition of their native tongue by attitudes towards that language or previous knowledge.

Under this model, the phonological perception of a speaker is in terms of phonemes and not surface phonetic segments; it is separated from the surface by the fewest number of substitutions. The learner's perception, therefore, is in terms of the phonemes selected to function in their native language. The more effectively they train themselves to perceive the phonetic realizations of L2-specific sound intentions, the sooner they will "decipher" the latter. Only then may a consistently correct production follow.

With the proposed NGTA, we aim to extend and modify the explanatory potential of natural phonology to account for the process of language acquisition from a multilingual perspective. In particular, we will apply a more rigid natural linguistic methodology to the phenomenon of multilingual acquisition. We will divide our evidence into system-internal and external, refer to universal preferences (i.e., preferability generalizations) rather than universals, allow for dynamic emergence rather than unsuppression, allow for both inductive and principled deductive explanations, include usage-based frequency-driven explanations, and expand the array of extralinguistic factors, including individual ones, which impact acquisition.

3.1.2 Insights from Complex Systems

A complex system is "a system in which large networks of components with no central control and simple rules of operation give rise to complex collective behaviour, sophisticated information processing, and adaptation via learning or evolution" (Mitchell, 2009, p. 13). Complex systems represent all those aspects of the natural world that do not conform to cause-and-effect, reductionist explanations. As such, they have been described and explored in physical, biological, and computer sciences, as well as in economics, linguistics, and particularly applied linguistics (see Hiver & Al-Hoorie, 2020).

According to Kretzschmar (2015, p.c.), applying complexity theory to linguistics constitutes a paradigm shift not yet fully appreciated in present-day research. Much of linguistics has relied on explanations deriving from independently established universals or on universal preferences based on general principles. In complex systems, universals are replaced by post-factum generalizations formulated on the basis of observed emergent effects, which have arisen randomly rather than being determined by a specific cause. Rather than sharing the same underlying system of language, "we all participate in speech, but the language is a little different… for each one of us individually… (R)ules and systems… are generalizations that we make after the fact from our perceptions" (Kretzschmar, 2015, p. 3). In speech treated as a complex system, every individual usage needs to be considered, much in line with usage-based approaches to phonology (Bybee, 2001, 2007, 2010). Such approach is grounded, among others, in Zipf's law (1949) (the law named after the American linguist George Kingsley Zipf) according to which, for example, words in texts rank roughly inversely proportional to frequency, that is, there are very few frequent words, some moderately frequent words, while most words occur at very low frequencies.

Kretzschmar (2015) is critical of generative and structural linguistics for their reductionist perspective. Natural linguistics, however, fulfils most of his expectations for a theory of language. In fact, self-organization was adopted by natural linguistics already in the early 1990's for the acquisition of morphology via proto- and pre-morphology to modularized morphology (Dressler & Peltzer-Karpf, 1995; Karpf, 1990) as well as for the acquisition of phonology via the stages of pre-phonology and proto-phonology to reach language-specific phonology (Dziubalska-Kołaczyk, 1997, 2002). Both in the acquisition of phonology and morphology, children's selection of input data was seen as guided by self-organizing processes. Children's preferences of selection were based on "saliency, frequency and repeated occurrence in a comparable configuration" (Dressler, 1997, p. 10). Importantly, those studies concerned the developmental stages in first language acquisition, while our proposed NGTA embraces multilingual acquisition of speech at any stage in life.

Classic natural phonology (Donegan & Stampe, 2009) is not incompatible with the emergentism of complex systems, either. As Donegan (1985) notes, "It would not alter the theory of natural phonology substantially to say that processes may be discovered by the child as he learns to use his vocal tract" (p. 26, note 5). Instead of saying that processes are discovered, one might say that they emerge as a result of the child's struggling to make efficient use of their inborn physical (articulatory and perceptual) abilities in order to overcome difficulties posed by this task. Processes emerge universally; this, however, does not imply that they are identical for all children. Since children are active in acquisition, and they are influenced by a particular ambient language, they discover divergent solutions to the difficulties they face, retreat from already entered paths, and so on.

A number of L2 researchers have examined complex systems. For example, Larsen-Freeman (1997) suggested complexity science for the study of language acquisition. Ellis and Larsen-Freeman (2009) discovered the nonlinear pattern for ESL in *Language as a Complex Adaptive System* (for the most recent review of research methods for complexity theory in applied linguistics and the Complex Dynamic Systems Theory, see Hiver & Al-Hoorie, 2020). Bybee (2001) discussed complex systems and their emergent properties claiming that Lindblom et al. (1984) were the first to apply the notion of emergent structure in linguistics. In Bybee's interpretation of complex systems, substance (i.e., phonetics and semantics) and use interact to create structure. The pillars of Bybee's well-established usage-based approach to language are frequency effects, creativity of repetition, and the notion of schemas or emergent generalizations.

We believe that the perspective of speech as a complex system has a potential to enlighten and enhance our proposal of NGTA. On the one hand, complex systems deal with dynamically changing states, which is compatible with the process of language acquisition. On the other hand, complex systems allow stable patterns to emerge—that is, "patterns on which we rely for effective communication in social interactions" (Kretzschmar, 2015, p. 20)—which guarantees language attainment. Crucial in the complex systems approach is the role of experience with frequent forms, and individual differences. Every distribution becomes non-linear in the sense that the most frequent forms (sounds, words) become preferred by an individual or in a corpus, and all those distributions get the same A-curve shape. In other words, there is a wide range of possible form realizations given by each and every individual, but some are heard more often than others, which leads to higher activation of selected neural pathways and to habituated patterns. Although the more frequent forms are used most often, we also need the less frequent forms in the tail of the distribution (for example, for a specialized terminology). Consequently, it is impossible to generalize over a large population or the whole language, since languages are locations at a grand continuum of speech (Kretzschmar, p.c.). This reasoning can be adapted to the situation of acquisition[1] of an additional language within the proposed theory of NGTA.

3.2 NGTA's Epistemological Stand

To arrive at a comprehensive account of multilingual acquisition, NGTA combines the apparently divergent epistemologies of two theories: natural phonology and complex systems (see Sect. 3.1). Thus, on the one hand, it embraces the communicative and cognitive orientation of language and the conditioning impact of extralinguistic factors within a functionalist perspective and, on the other hand, it includes a non-teleological perspective.

As Fig. 2 shows, NGTA is eclectic in proposing both principled explanations

[1] We propose the term situation of acquisition to embrace all aspects of a given acquisition case (be it of L1 or Ln, by an individual or a population, in a formal or natural context, at a given age, given proficiency level, etc.).

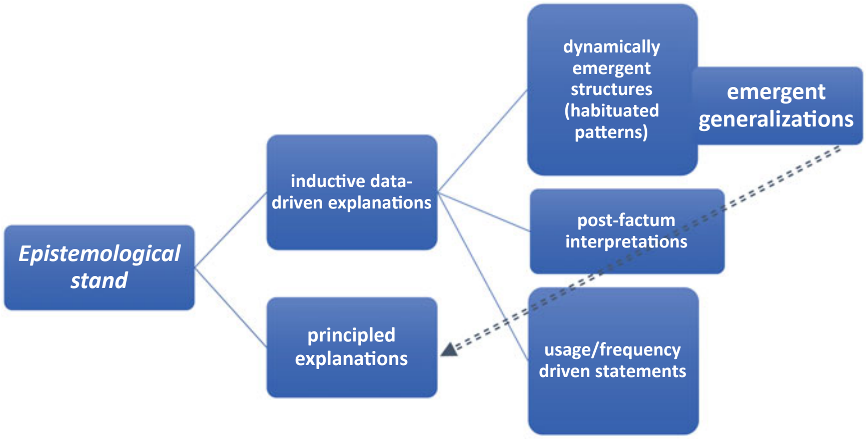

Fig. 2 NGTA's epistemological stand

and inductive data-driven accounts. The latter subsume post-factum interpretations, usage-based frequency driven statements, and dynamically emergent structures, also referred to as habituated patterns. To make our epistemological stand clear: it would not alter our theory if the principled explanations were, in fact, emergent generalizations which at some point in the history of science had given rise to the formulation of principles. In the present state of knowledge, these explanations already exist as in the extralinguistic principles (e.g., semiotic, functional, and cognitive, as specified in Sect. 3.1.1) while other regularities are yet emerging.

Figure 2 summarizes our epistemological stand. The post-factum interpretations receive support from the usage-based approach (Bybee, 2001), complexity theory and typological research on language universals. As Haspelmath (2016) explains, functional-adaptive explanations, which make reference to language use and language change, are necessarily explanations of language universals; they do not make unwarranted (aprioristic) uniformity assumptions; further, they are readily testable by cross-linguistic and usage data. In fact, Haspelmath questions whether it matters where universals come from. Since it is naïve and unfeasible to expect to investigate all languages data, we generalize on the basis of what we have access to as well as we look for explanations outside of language. Hence, in NGTA, we propose to use the term *preferability generalizations* (noted as emergent generalizations in Fig. 2) instead of universals.

NGTA aims to embrace the complexity of the process of multilingual acquisition in its entirety. Therefore, it takes into account both linguistic and extralinguistic variables as potential sources of influence and moderating factors (see Fig. 3).

The linguistic variables can be classified into:

1. L1—the first language.
2. Ln—languages other than L1 (including L2, L3, and further languages; their language specific systems and typological properties).

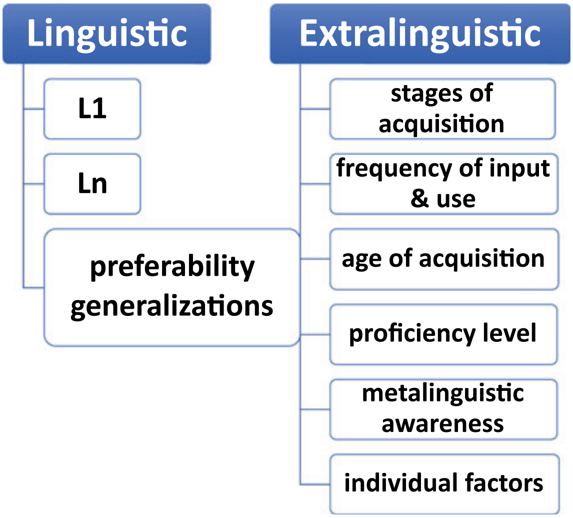

Fig. 3 Linguistic and extralinguistic variables in NGTA

3. Preferability generalizations (both deductive and inductive).

The proposed list of extralinguistic variables includes:

1. Stages of acquisition (i.e., a continuum from initial to advanced or proficient, order of acquisition, recency of use).
2. Frequency of input and use (i.e., exposure, amount of training, amount of language use).
3. Age of acquisition (i.e., the younger the speaker, the less complex the network of interdependencies).
4. Proficiency level (i.e., attainment in particular languages shapes dominance relations and cross-linguistic influence—CLI).
5. Metalinguistic awareness.
6. Individual factors (e.g., attitude, motivation, aptitude, personality, self-evaluation, cognitive factors, such as attention and memory).

3.3 NGTA's Assumptions

On the basis of the analysis of the network of interdependencies formed by the linguistic and extralinguistic variables described in Sect. 3.2, we formulated three general assumptions about the process of multilingual acquisition:

> GA I. All three linguistic variables (L1, Ln, preferability generalizations) may influence the process, that is, none can be excluded. However, the relative impact of each variable is moderated by the configuration of extralinguistic factors in a given acquisition situation.

> GA II. Complexity of the acquisition process is dynamic and grows as a function of time and language learning experience, that is, the older the multilingual learners are, the more

complex the network of interdependencies between and among linguistic and extralinguistic variables becomes, which, in turn, may facilitate or inhibit acquisition. In this sense, age of acquisition is negatively related to the complexity of the network of interdependencies among the variables.

GA III. Motivated by a distinction proposed by Kahneman (2011), NGTA distinguishes two levels in the language acquisition process. Level 1 is automatic, in the sense of involuntary and instinctive, as manifested by, for example, articulatory routines and phonetic perceptual constraints. Level 2 is conscious, cognitively-based, mindful, cognizant and metalinguistic, as manifested by any aspect of meta-awareness. It is a meta-level of acquisition. At Level 1, it is the L1 that prevails as the source of CLI in acquisition, while at Level 2 other Ls surface as CLI sources.

Summing up NGTA's general assumptions, all variables matter, complexity grows with age, and acquisition proceeds at the automatic and metalinguistic levels.

In NGTA, growth of attainment in the additional language is regulated by the degree of complexity of interdependencies among the relevant variables. Phonology grows in a learner along his or her individual natural path of acquisition. All learners are influenced by linguistic and extralinguistic variables (Fig. 3). The fundamental research question is how much the individual paths converge and to what extent they remain divergent.

4 NGTA—Data Support from (Mor)Phonotactics

In this section, we want to juxtapose the general assumptions of NGTA and formulate specific hypotheses for the data obtained from research on phonotactics and morphonotactics. Phonotactic grammar is concerned with well-formedness of consonant clusters and operates on basic, non-derived, lexical forms (e.g., the final clusters in *band* and *past*). Morphonotactics takes care of the remaining, morphologically complex, forms (e.g., the final clusters in *ban(n)* + *ed* and *pass* + *ed*). Morphonotactics is the area of interaction between morphotactics and phonotactics (Dressler & Dziubalska-Kołaczyk, 2006) and shows how inflection, word-formation and compounding contribute to the creation of consonant clusters. In our studies, we demonstrate that the interaction between phonotactics and morphonotactics provides a richer insight into the understanding of cluster complexity. Crucially, one expects relatively marked clusters across morpheme boundaries and relatively unmarked ones within morphemes. This expectation has direct consequences for the acquisition of clusters.

Guided by the general assumptions presented in Sect. 3.3, and informed by the data generated in our research on (mor)phonotactics (Dziubalska-Kołaczyk & Zielińska, 2010, 2011; Dziubalska-Kołaczyk & Zydorowicz, 2014; Marecka & Dziubalska-Kołaczyk, 2014), we formulated the following hypotheses pertaining to the process of multilingual acquisition of clusters:

H.1. Universal phonotactic preferences (= preferability generalizations) influence the acquisition of consonant clusters in a second/foreign language.

H.2. The hierarchy of the universal phonotactic preferability will correlate with the level of difficulty in pronouncing L2/Ln lexical clusters.

H.3. Morphonotactic clusters carry morphological information, and their markedness is used to signal their function.

H.4. L2/Ln learners will put some effort into the acquisition of morphonotactic clusters, despite their phonotactically dispreferred status.

More specifically, we predicted the following for L2/FL consonantal phonotactics and morphonotactics:

1. Clusters are difficult.
2. Clusters common across languages are easier.
3. Preferred (unmarked) clusters are easier.
4. Shorter clusters are easier.
5. Clusters are acquired in this order: medial > initial > final.
6. Dispreferred (marked) clusters are difficult also when they are morphonotactic.
7. Children may learn morphonotactic clusters earlier.
8. Frequent (token frequency) clusters are easier.
9. Proficiency and metalinguistic awareness enhance the learning of clusters.

Figure 4 displays a summary of our predictions based on the linguistic and extralinguistic variables of NGTA.

We conducted four studies to check our hypotheses. In the first one (Dziubalska-Kołaczyk & Zielińska, 2010), 53 young learners of L2 English (between 11 and

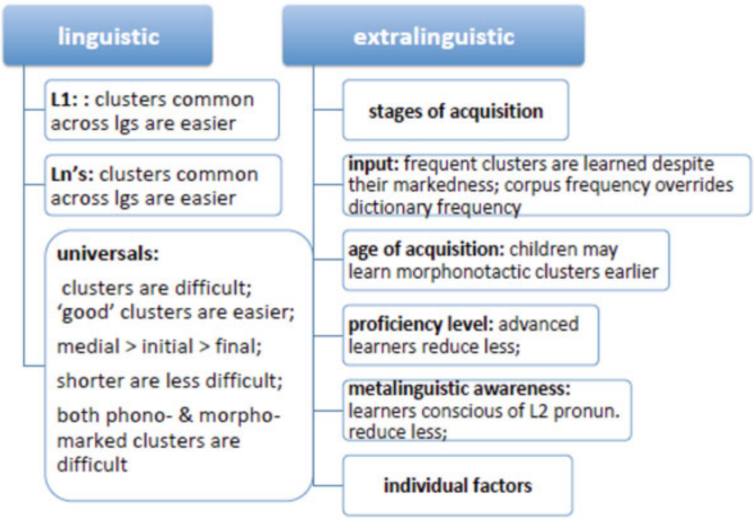

Fig. 4 (Mor)Phonotactic predictions for the acquisition of consonant clusters in light of linguistic and extralinguistic variables of NGTA

13 years old) read a carrier phrase with 83 nonce words containing a cluster. They were native speakers of the following 15 languages: Japanese, Korean, Vietnamese (independent), Chinese (Sino-Tibetan), Kosraean, Marshallese, Palauan, Ponapean, Samoan, Tagalog, Trukese, Visayan (Austronesian), Tamil (Dravidian), and Polish (Slavic). Word-level results showed that the learners produced medial clusters more successfully than initial and final clusters (with 58%, 70%, and 76% incorrect renditions, respectively). Further, it was easier for them to produce less complex clusters than more complex ones (i.e., 36% of 2-consonant clusters and 10% of 3-consonant clusters were correctly produced). Correlations were found for both initial and final clusters between their degree of preferability ("goodness") and the proportional degree of difficulty in their production (the number of errors). The less preferred a cluster was, the more difficult it appeared, with the strongest correlation showing for the degree of reduction of a difficult cluster.

In the second study (Dziubalska-Kołaczyk & Zielińska, 2011), 16 Asian learners of Polish (8 Koreans, 7 Chinese, 1 Thai), aged between 19 and 40 years old, with an average length of learning Polish of 1.5 years, read a carrier phrase with 40 words containing a double cluster, in all word positions, in both lexical and morphonotactic contexts. The participants had substantially more problems with morphonotactics than phonotactics. This finding may suggest that predictions about cluster acquisition should differ with reference to L1 and L2. Whereas in L1 acquisition the difference between lexical and morphonotactic clusters is important from the start and morphonotactics is acquired earlier because it is more salient and useful, in L2 both types of clusters need to be learned so their status is leveled out—marked clusters are more difficult for the learner no matter their status. Interestingly, the participants turned out to be too advanced and too conscious of their pronunciation in Polish, pointing to the role of proficiency and metalinguistic awareness as possibly conditioning their performance.

In the third study (Dziubalska-Kołaczyk & Zydorowicz, 2014), 10 subjects participated, all L1 speakers of CV languages (Chinese, Japanese, and Korean), aged 21–47, who had learned Polish for 2–5 years, with length of residence in Poland from 2 months to 6 years. The control group consisted of three native speakers of Polish whose mean age was 26.6. The task was to read aloud a text in a casual tempo with 58 clusters embedded in 117 words in all positions. The clusters were selected from a corpus (which was an advantage over the previous study). We tested three hypotheses derived from the *Net Auditory Distance (NAD) Principle*—that is, a measure of phonotactic goodness of clusters based on auditory distances between consonants in a cluster resulting from such phonetic features as the manner of articulation (MOA), place of articulation (POA) and voicing (for a detailed discussion, see Dziubalska-Kołaczyk, 2014). The three hypotheses were (a) the number of cluster simplification processes would increase along with cluster length; (b) word-final position consonant clusters would be the most vulnerable to change, with word-initial being the most salient, and word-medial being the most tolerant of all the consonant clusters; and (c) preferred consonant clusters would be preserved in L2 production more successfully than dispreferred consonant clusters. The study yielded the following results. First, shorter clusters were produced more accurately than longer sequences. Second, double, triple and quadruple clusters in word-final position turned out to be the most susceptible to simplification processes. As regards the "safety" of medials, double medials turned out to be modified more frequently than initials, whereas triple medials were simplified as frequently as triple initials. Only in the group of quadruples were medials more stable than initials and finals. Double initials posed the least difficulty for second language learners, probably due to the salience of the initial position, which carries high informational load. Finally, with reference to cluster markedness, dispreferred consonant clusters underwent simplification more frequently than preferred consonant clusters; for triple medials the reduction rates were equal to preferred and dispreferred clusters (i.e., the phonological preference was overridden by the word position itself). The overall cluster modification rate amounted to 34%, which pointed to the fact that clusters posed a challenge to L2 learners of Polish. Another potential criterion which may have had an impact on cluster production was the morphological composition of a cluster. Although the sample contained clusters which were purely phonotactic, purely morphonotactic and mixed, reliable conclusions could not be drawn: the non-native participants considered the text challenging and admitted that some words were unfamiliar to them. Indeed, word familiarity is considered a prerequisite for the study of the influence of morphology on cluster status.

In the fourth study (Marecka & Dziubalska-Kołaczyk, 2014), we investigated 50 Polish kindergarten children, who produced six initial sC clusters: /st/, /sp/, /sk/, /sx/, /sm/ and /sw/ elicited in a picture naming task. The sC clusters were selected in such a manner that they were similar in terms of structure, yet differed with respect to preferability. In accordance with the applied NAD principle (see Dziubalska-Kołaczyk, 2014), we predicted that the /st/ would be the most dispreferred (reduced) cluster, followed by /sp/, /sk/ and /sx/, /sm/ and finally /sw/. We calculated the ratio of clusters in which the consonants were distorted, substituted or reduced to all the clusters produced; however, the results only partially confirmed our predictions. Contrary to the NAD index, a strong, statistically significant preference for the /st/ cluster was observed with children producing frequent structures with greater accuracy than less frequent ones. On the other hand, /sx/ was the cluster with the highest rate of mispronunciations. It follows that corpus frequency of the structures must be taken into consideration, although it does not fully account for the results since it is both corpus frequency data and acquisition data that are shaped by universal principles. Further, morphonotactic clusters, whose function is to mark morphological boundaries, have a higher markedness status than the phonological (lexical) clusters, thus the former are found to be more difficult for children to produce.

In conclusion, our studies supported the following predictions for L2/FL consonantal phonotactics and morphonotactics:

1. Clusters are difficult for learners so they modify and reduce them (study 1, 2, 3, 4).
2. Difficulty correlates with the universal phonotactic preferences: 'good' clusters are easier (study 1, 2, 3).
3. Order of 'survival' of clusters in a word is: medial > initial > final (study 2, 3).

4. Less complex clusters (shorter) are less difficult (study 1, 2, 3).
5. Proficiency and metalinguistic awareness enhance the learning of clusters (study 3).

Further extending NGTA's predictions, our studies also showed the following:

1. Marked clusters are difficult for learners no matter whether they are phonotactic or morphonotactic (while children may learn some morphonotactic clusters earlier) (study 4).
2. Frequently used clusters are learned despite their markedness (corpus frequency overrides dictionary frequency) (study 4).
3. Cluster types which are common across languages (e.g., st-) seem to be easier despite their markedness, based on general observations (study 4).

All in all, the four studies summarized above allowed us to corroborate some of the predictions formulated within the framework of NGTA, but not all, therefore further work is still needed to reinforce and/or verify the findings.

5 Conclusions and Implications

Our goal was to put forward a new theory of language acquisition of speech and support it with empirical evidence from (mor)phonotactics. NGTA is holistic in the sense that it incorporates each and every aspect of the acquisition process. One of its main assumptions is a gradual dynamic emergence of Ln phonology, shaped by input from the L1 and other Ls, and influenced by typology, universal preferences (in the sense of preferability generalizations) and context. NGTA qualifies the impact of input on the basis of the type and token frequencies of its elements and considers the preference-based, typological and language-specific aspects of the growth. The proposed Natural Growth Theory of Acquisition has got implications for modeling the acquisition of foreign language speech, and may, indirectly, have impact on pronunciation teaching and learning. The findings may assist in overcoming difficulties in teaching complex phonotactics as well as pronunciation features such as VOT or vowel quality and duration, the realization of which may differ across languages being acquired in the multilingual context. Moreover, the implications of the model may be applied to raising metalinguistic awareness concerning, in general, foreign language phonology, and, specifically, the phenomenon of foreign accentedness in L2/L3 speech. All in all, the modelling potential offered by NGTA is relevant for any aspect of teaching foreign language speech, as well as for understanding the individual variation that is commonly attested in multilingual speech acquisition.

Recapitulating, NGTA is conceived as a general theory of language acquisition, which allows us to explain the acquisition of morphology, phonology and other language domains. For instance, our research in progress focuses on the reanalysis of L3 phonological data from studies by Wrembel (2015) with the view to interpreting them as data support for NGTA. In future research we plan to expand analyses to further domains (e.g., syntax) in order to corroborate NGTA's predictions and confirm its explanatory potential as a comprehensive theory accounting for the acquisition of language from a multilingual perspective.

References

Best, C. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171–204). York Press.

Best, C., & Tyler, M. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning* (pp. 13–34). John Benjamins. https://doi.org/10.1075/lllt.17.07bes

Burkette, A., & Kretzschmar, W., Jr. (2018). *Exploring linguistic science: Language use, complexity, and interaction*. Cambridge University Press. https://doi.org/10.1017/9781108344326

Bybee, J. (2001). *Phonology and language use*. Cambridge University Press. https://doi.org/10.1017/CBO9780511612886

Bybee, J. (2007). *Frequency of use and the organization of language*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195301571.001.0001

Bybee, J. (2010). *Language, usage and cognition*. Cambridge University Press. https://doi.org/10.1017/CBO9780511750526

Chang, C. B. (2019). The phonetics of second language learning and bilingualism. In W. F. Katz & P. F. Assmann (Eds.), *The Routledge handbook of phonetics* (pp. 427–447). Routledge. https://doi.org/10.4324/9780429056253-16

Donegan, P. J. (1985). "How learnable is phonology?" In W. Dressler, & L. Tonelli (eds.), *Papers on natural phonology from Eisenstadt. Padova, Italy:CLESP (Cooperativa Libraria Editoriale Studentesca Patavina)*. 19–31.

Donegan, P. J., & Stampe, D. (2009). Hypotheses of natural phonology. *Poznan Studies in Contemporary Linguistics, 45*(1), 3–31. https://doi.org/10.2478/v10010-009-0002-x

Dressler, W. U. (1984). Explaining natural phonology. *Phonology Yearbook, 1*, 29–50. https://doi.org/10.1017/S0952675700000282

Dressler, W. U. (1996). Principles of naturalness in phonology and across components. In B. Hurch & R. Rhodes (Eds.), *Natural phonology: The state of the art* (pp. 41–52). De Gruyter Mouton. https://doi.org/10.1515/9783110908992.41

Dressler, W. U. (1997). Introduction. *Papers and Studies in Contrastive Linguistics, 33*, 9–14. https://doi.org/10.1515/9783110886733.1

Dressler, W. U., & Dziubalska-Kołaczyk, K. (2006). Proposing morphonotactics. *Rivista di Linguistica*, *18*(2), 249–266.

Dressler, W. U., & Peltzer-Karpf, A. (1995). The theoretical relevance of pre-and protomorphology in language acquisition. *Yearbook of Morphology, 1994*, 99–122. https://doi.org/10.1007/978-94-017-3714-2_4

Dziubalska-Kołaczyk, K. (1990). *A theory of second language acquisition within the framework of natural phonology*. Adam Mickiewicz University Press.

Dziubalska-Kołaczyk, K. (1997). Pre- and proto- in Polish phonology and morphology and their interrelations. *Papers and Studies in Contrastive Linguistics, 33*, 159–171. http://ifa.amu.edu.pl/psicl/files/33/13Dziubalska-Kolaczyk.pdf

Dziubalska-Kołaczyk, K. (1998). Self-organization in early phonology. In S. Puppel (Ed.), *Scripta Manent. Publikacja wydziałowa z okazji 10-lecia istnienia Wydziału Neofilologii UAM 1988–1998* (pp. 99–112). Motivex.

Dziubalska-Kołaczyk, K. (2002). *Beats-and-binding phonology*. Peter Lang. https://www.peterlang.com/document/1094101

Dziubalska-Kołaczyk, K. (2009). NP extension: B&B phonotactics. *Poznań Studies in Contemporary Linguistics, 45*(1), 55–71. https://doi.org/10.2478/v10010-009-0011-9

Dziubalska-Kołaczyk, K. (2012). Modern natural phonology and phonetics. In E. Cyran, H. Kardela, & B. Szymanek (Eds.), *Sound, structure and sense: Studies in memory of Edmund Gussmann* (pp. 199–210). Wydawnictwo KUL.

Dziubalska-Kołaczyk, K. (2014). Explaining phonotactics using NAD. *Language Sciences, 46*, 6–17. https://doi.org/10.1016/j.langsci.2014.06.003

Dziubalska-Kołaczyk, K. (2019, August 27–30). *Natural growth theory of acquisition (NGTA): Evidence from (mor)phonotactics* [Conference presentation]. 3rd ISMBS 2019, Chania, Crete, Greece.

Dziubalska-Kołaczyk, K., & Zielińska, D. (2010). Predicting phonotactic difficulty in second language acquisition. In A. S. Rauber, M. A. Watkins, R. Silveira, & R. D. Koerich (Eds.), *The acquisition of second language speech: Studies in honor of Professor Barbara O. Baptista* (pp. 281–304). Editora Insular.

Dziubalska-Kołaczyk, K., & Zielińska, D. (2011). Universal phonotactic and morphonotactic preferences in second language acquisition. In K. Dziubalska-Kołaczyk, M. Wrembel, & M. Kul (Eds.), *Achievements and perspectives in SLA of speech: New sounds 2010* (pp. 53–64). Peter Lang.

Dziubalska-Kołaczyk, K., & Zydorowicz, P. (2014). The production of high-frequency clusters by native and non-native users of Polish. In *Proceedings of the International Symposium on the Acquisition of Second Language Speech, Concordia working papers in applied linguistics,* 5, COPAL (pp. 130–144). http://doe.concordia.ca/copal/documents/11_Dziubalska-Kolaczyk_Zydorowicz_Vol5.pdf

Dziubalska-Kołaczyk, K., & Wrembel, M. (2017, September 10–13). *Natural growth model: Explaining third language phonological acquisition* [Conference presentation]. 50th annual meeting of the Societas Linguistica Europaea (SLE 2017), Zürich.

Ellis, N., & Larsen-Freeman, D. (Eds.). (2009). *Language as a complex adaptive system.* Wiley-Blackwell.

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. https://doi.org/10.1016/S0095-4470(19)30537-6

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. 233–272). York Press.

Gut, U. (2010). Cross-linguistic influence in L3 phonological acquisition. *International Journal of Multilingualism*, *7*, 19–38.

Gut, U., Fuchs, R., & Wunder, E. M. (Eds.). (2015). *Universal or diverse paths to English phonology.* Berlin: Mouton De Gruyter.

Haspelmath, M. (2016, August 31–September 3). *Explaining universals of syntactic argument coding* [Plenary speech]. 49th annual meeting of the Societas Linguistica Europaea (SLE 2016), Naples.

Hiver, P., & Al-Hoorie, A. H. (2020). *Research methods for complexity theory in applied linguistics.* Multilingual Matters.

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar.

Karpf, A. (1990). *Selbstorganisationsprozesse in der sprachlichen Ontogenese.* Narr.

Kretzschmar, W. A., Jr. (2015). *Language and complex systems.* Cambridge University Press. https://doi.org/10.1017/CBO9781316179017

Kuhl, P. K., & 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). York Press.

Larsen-Freeman, D. (1997). Chaos/complexity science and second language acquisition. *Applied Linguistics, 18*(2), 141–165. https://doi.org/10.1093/applin/18.2.141

Leather, J., & James, A. (1991). The acquisition of second language speech. *Studies in Second Language Acquisition, 13*(3), 305–341. https://doi.org/10.1017/S0272263100010019

Lindblom, B., MacNeilage, P., & Studdert-Kennedy, M. (1984). Self-organizing processes and the explanation of language universals. In B. Butterworth, B. Comrie, & Ö. Dahl (Eds.), *Explanations for language universals* (pp. 181–203). Berlin: Walter de Gruyter & Co.

Marecka, M., & Dziubalska-Kołaczyk, K. (2014). Evaluating models of phonotactic constraints on the basis of sC cluster acquisition data. *Language Sciences, 46*, 37–47. https://doi.org/10.1016/j.langsci.2014.06.002

Mitchell, M. (2009). *Complexity: A guided tour*. Oxford University Press.

Pisoni, D. B. (1996). Some thoughts on "normalization" in speech perception. In K. Johnson & J. W. Mullennix (Eds.), *Talker variability in speech processing*. Academic Press.

Rubin, E. (1915). *Synsoplevede Figurer*. Glydendalske Boghandel.

Stampe, D. (1969). *The acquisition of phonetic representation*. Indiana University Linguistics Club.

Stampe, D. (1979). *A dissertation on natural phonology*. Garland.

Strange, W. (1999). Levels of abstraction in characterizing cross-language phonetic similarity. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. C. Bailey (Eds.), *Proceedings of the 14th International Congress of Phonetic Sciences* (pp. 2513–2519). International Phonetic Association. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS1999/p14_2513.html

Wrembel, M. (2012). Foreign accentedness in Third Language Acquisition: The case of L3 English. In J. Cabrelli Amaro, S. Flynn, & J. Rothman (Eds.), *Third language acquisition in adulthood* (pp. 281–309). Amsterdam: John Benjamins.

Wrembel, M. (2015). *In search of a new perspective: Cross-linguistic influence in the acquisition of third language phonology*. Wydawnictwo Naukowe UAM. https://press.amu.edu.pl/pl/in-search-of-a-new-perspective-crosslinguistic-influence-in-the-acquisition-of-third-language-phonology-3495.html

Wrembel, M., & Dziubalska-Kołaczyk, K. (2016, August 31–September 3). *Explaining third language phonological acquisition from the Natural Phonology perspective* [Conference presentation]. 49th annual meeting of the Societas Linguistica Europaea (SLE 2016), Naples.

Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Addison-Wesley Press.

Katarzyna Dziubalska-Kołaczyk is Full Professor and Vice-Rector of Adam Mickiewicz University in Poznań, Poland. She was Dean of the Faculty of English (2012–2019), Senior Fulbright scholar (2001–2002) and President of Societas Linguistica Europaea (2013–2014). She has published extensively on phonology, phonetics, and language acquisition. Her books include A Theory of Second Language Acquisition within the Framework of Natural Phonology, Beats-and-Binding Phonology and Phonotactics and Morphonotactics of Polish and English. She edits Poznań Studies in Contemporary Linguistics and organizes The Poznań Linguistic Meetings (annual conference).

Magdalena Wrembel, Ph.D., D. Litt., is University Professor at the Faculty of English, Adam Mickiewicz University in Poznań, Poland. Her main research areas involve phonetics and phonology, SLA of speech, third language acquisition, multilingualism, and language awareness as well as innovative trends in pronunciation pedagogy. She has published extensively in international journals and edited collections. She is Deputy Head of Bilingualism Matters@Poznań.

Perceptual Drift in L1 Phonetic Categories in Multilinguals

Jolanta Sypiańska and Zuzanna Cal

Abstract The aim of this chapter is to investigate L1 perceptual drift in late-onset multilingual learners. L3 as the source of L1 drift was isolated by examining the open-mid front unrounded vowel /ε/ in 36 speakers of L1 Polish, L2 English, and L3 Spanish with a comparable proficiency in their L2 and L3. The mid front vowel is close-mid in Spanish but open-mid in the other two languages in the linguistic repertoire of the speakers. There were three groups of participants based on their L3 (Spanish) exposure: Instruction group (Polish residents exposed to Spanish through instruction; $n = 14$), Immersion group (Polish immigrants to Spain exposed to Spanish through immersion; $n = 10$), and Control group (no exposure to Spanish; $n = 12$). A 685 Hz–640 Hz vowel continuum was synthesised and correctness in distinguishing between cross-boundary pairs of vowels was assessed using a discriminatory AX task. The results revealed greater L1 perceptual drift in the Immersion group than in the Instruction and Control groups. The Immersion group experienced a boundary shift towards a higher F1 value (i.e., a lower L1 vowel), which we understand to constitute dissimilatory L1 drift (away from the L3 source vowel) for the purpose of category differentiation.

Keywords Perceptual L1 drift · Dissimilatory L1 drift · Discriminatory task · Phonetic category · Language exposure

---

J. Sypiańska (✉)
University of Szczecin, Szczecin, Poland
e-mail: jolanta.sypianska@usz.edu.pl

Z. Cal
Adam Mickiewicz University, Poznań, Poland
e-mail: zuzanna.cal@amu.edu.pl

V. G. Sardegna and A. Jarosz (eds.), *Theoretical and Practical Developments in English Speech Assessment, Research, and Training*, Second Language Learning and Teaching,
https://doi.org/10.1007/978-3-030-98218-8_17

1 Introduction

Changes in the phonetic system of a healthy adult's native language may result from foreign language acquisition in the form of cross-linguistic influence from the foreign language(s) to the first language (L1) referred to as L1 drift (Chang, 2010). These changes have been reported in production (Chang, 2011, 2012; Schwartz & Wojtkowiak, 2017; Sypiańska, 2016, 2017) and less frequently in perception (Dmitrieva, 2010, 2019; Namjoshi et al., 2015; Tice & Woodley, 2012). Thanks to a series of studies on early-onset changes in the L1 due to second language (L2) exposure (Chang, 2010, 2011, 2012) and on long-term L1 phonetic attrition of immigrants residing in the L2 country (de Leeuw et al., 2011, 2017; Flege & Eefting, 1987; Major, 1992), L1 drift is now widely accepted as a phenomenon that takes place as part of the process of acquiring a foreign language. Yet, the conditions in which it is likely to occur are still not entirely clear. Also, there seems to be a paucity of studies on Lx learners who use their foreign language(s) extensively but remain in the L1 country. The purpose of this chapter is to investigate previously unexplored conditions for L1 drift as we look into L1 perceptual drift in multilingual speakers with different types of exposure to Lx.

The layout of the chapter is as follows. We provide a review of the literature on L1 drift with a particular focus on L1 perceptual drift. We then proceed to describing the design of the study and analysing the results. The implications of the study are then highlighted and further avenues of research are delineated. This is followed by the most important conclusions of the chapter.

2 Literature Review

2.1 L1 Drift

Changes in the phonetics of a healthy adult's native language may result from foreign language acquisition in the form of cross-linguistic influence from the foreign language(s) to the L1, that is to say, regressive transfer. These changes have been reported in production (Chang, 2011, 2012; Sypiańska, 2016, 2017) and perception (Namjoshi et al., 2015; Tice & Woodley, 2012). In Sancier and Fowler's (1997) seminal study on gestural drift, the authors describe a case of "perceptually guided changes in speech production" (p. 421) of the native language in an L1 Brazilian Portuguese, L2 English speaker. Notably, the VOT in the Brazilian Portuguese voiceless stops lengthened after a stay in the USA and shortened after a stay in Brazil as an effect of influence from the longer and then the shorter VOT of respectively English and Brazilian Portuguese stops. Early-onset changes are also reported in Chang (2010) in which the F1 of L1 English vowels rose after three weeks of an intensive L2 Korean course. His results also allowed him to observe a lengthening of the VOT in the English voiceless stops reflecting the longer L2 Korean VOT after only two weeks of exposure to Korean.

Bearing in mind that “L1 drift is a common but not necessarily inevitable part of the process” of Lx learning (Schuhmann & Huffman, 2015, abstract), the conditions in which L1 drift may be expected are still largely unverified. Chang (2019) distinguishes between phonetic drift and phonetic or phonological attrition. Phonetic drift is defined as “short-term L1 changes both in early- and late-onset L2 learners which are attributable to recent L2 experience” (p. 192). Phonetic attrition is viewed as “long-term L1 changes in late-onset L2 learners which are unlikely to be due to recent L2 experience only” (p. 192). In studies on production, Schwartz and Wojtkowiak (2017) found phonetic drift in the form of shorter prevoicing in L1 Polish voiced stops among Polish learners of L2 English when compared to Polish monolinguals. Schwartz et al. (2020) showed further support for L1 drift in the voiced stop series of advanced L1 Polish speakers of L2 English. Sypiańska (2021) observed L1 drift in the voiceless series of senior Polish learners of L2 English which she attributed to a greater mastery of aspiration in voiceless stops than prevoicing in prevoiced English stops. Also, Herd et al. (2015) described how L1 English, L2 Spanish speakers residing in the L1 country produced more peripheral vowels and voiced stops with greater negative VOTs in the L1 as a result of L2 influence. They concluded that L2 instruction in an L1-dominant environment can be conducive to phonetic drift in the L1 similarly to an L2 immersion setting.

When it comes to perception, Tice and Woodley (2012) found a lower voiced-voiceless boundary on a /pa-ba/ continuum in novice L1 English learners of L2 French who took part in a six-week study abroad course when compared to English controls. Their results confirmed the possibility of early-onset changes to the L1 in perception. In another study on perception, Dmitrieva (2019) showed evidence of differences in cue-weighting for stop voicing between L1 Russian, L2 English speakers residing in the USA and two control groups: monolingual Russian speakers residing in Russia and monolingual English speakers residing in the USA. The cue weighting for stop voicing depended on the language mode. In the Russian mode, L1 Russian speakers paid more attention to vowel duration and less to glottal pulsing in comparison to the Russian monolingual group. In Dmitrieva (2010), L1 Russian speakers of English residing in their L1 country used more ways to signal differences between word-final voiced and voiceless obstruents, including vowel duration and duration of voicing into closure, than Russian monolinguals. Finally, Namjoshi et al. (2015) reported different use of prosodic cues that signal word boundaries in the native language by L1 French speakers residing in the USA when compared to French speakers in the L1 country. They concluded that listeners have the ability of “tuning into the prosody of their linguistic environment” (abstract).

The cases discussed so far provide examples of assimilatory drift—that is, the L1 characteristic approximates the corresponding L2 element. However, L1 drift may also lead to dissimilatory drift—that is, a greater differentiation between the L1 and the L2 category. According to Flege (2007), dissimilation of phonetic categories may take place in more advanced L2 learners as a means of category differentiation. Schuhmann and Huffmann (2017) found evidence of L1 dissimilatory drift in L1

English learners of L2 Japanese. L1 dissimilation was present in the second formant of L1 English [ɑ] due to the influence of the L2 Japanese [a]. Still most speakers tended to assimilate the native vowel to the L2 counterpart especially in terms of F1. Their data suggest that the type of L1 drift is also speaker-specific and/or phonetic property-specific.

The scope of L1 drift research available in the literature seems to be to a great extent limited to speech production and L2 immersion settings. When compared, it is the L2 immersion setting which triggers greater L1 drift if any (Dmitrieva, 2019; Lang & Davidson, 2017). Furthermore, if L1 drift in production is perceptually guided due to Lx exposure, it will be interesting to find further evidence of L1 perceptual drift. When it comes to the setting, it would be vital to see whether late-onset multilingual learners remaining in the L1 country with an ongoing exposure to their foreign languages via formal instruction also experience L1 perceptual drift.

2.2 Vowel Perception

Although consonant perception, especially of stops, is unequivocally stated to be categorical, whether vowels are perceived categorically or continuous is a subject of debate. The continuous perception of vowels seems to be widely accepted in early literature (e.g., Fry et al., 1962; Stevens et al., 1969). It is explained that the existing discrepancies in coding between stops and vowels, such as differences in acoustic cues, may lead to the two kinds of sounds being perceived in an unlike manner. However, newer studies on vowel perception (e.g., Pisoni & Lazarus, 1974; Zhang & Shi, 2014; Zhang et al., 2016) suggest otherwise. For example, Zhang and Shi (2014) found evidence for categorical perception of Mandarin vowels along the /a/-/u/ continuum. The study pointed to the discrimination peak being located in the categorical boundary. Thus, it might be assumed that categorisation of vowels exists but is considerably weaker than that of (stop) consonants and continuous perception prevails over it.

Categorical perception research takes into consideration the relation between discrimination and identification of a continuum between two sounds. In one of the first studies on this matter, Liberman et al. (1957) speculate that identification limits discrimination of speech sounds in a sense that two different sounds can only be discriminated if they are identified differently. Hence, studies aimed at investigating categorical perception of sounds include identification and/or discrimination tasks. The former is based on labelling certain sounds based on whether they belong to the same category. The latter is based on whether two sounds are perceived the same or different. It is said to be easier to discriminate sounds across a category boundary than within a category (Boersma & Chládková, 2010). The most popular discrimination tasks used in research on categorical perception of speech include but are not limited to same-different (AX), matching-to-sample (ABX), two alternative forced choice (2AFC), and 4-interval forced choice (4IAX) (for a more extensive review, see Gerrits & Schouten, 2004).

3 The Study

The general aim of the study is to investigate L1 perceptual drift in late-onset multilingual learners. In particular, the study seeks to analyse the perception of the open-mid front unrounded vowel /ɛ/ by L1 Polish, L2 English and L3 Spanish speakers with a comparable proficiency in their L2 and L3 but different type and amount of exposure to the latter.

First of all, we aim to show an influence from the L3 on the L1 whose source can be isolated by means of looking at the mid front vowel which is high in Spanish (i.e., a close /e/) but low in Polish and English (i.e., an open /ɛ/) in the linguistic repertoire of the speakers. Perceiving a higher mid front vowel in their L1 Polish in comparison to a control group (i.e., L1 Polish, L2 English speakers with no exposure to Spanish) should be an effect of L1 drift whose origin may be traced to the influence of the L3. We further aim to investigate the effect of context of learning which contributes to different degrees of exposure to the foreign language on the amount of L1 drift. We foresee a greater amount of L1 drift in an immersion setting than in a formal instruction setting because exposure to the source language (L3 Spanish) is greater in the latter. A final aim is to determine whether the different contexts of learning and using the L3 may lead to different directions of L1 drift. Less Lx exposure (formal instruction) is hypothesised to evoke assimilatory L1 drift, and more exposure (Lx immersion setting) should favour dissimilatory L1 drift.

Thus, with a focus on the open-mid front unrounded vowel /ɛ/, the following research questions guided our investigation:

1. To what extent does the L3 in late-onset multilingual learners affect their perception of L1 phonetic categories (perceptual L1 drift)?
2. What is the effect of context of learning and using the L3 on the amount of perceptual L1 drift?
3. Do different contexts of learning and using the L3 contribute to different directions of perceptual L1 drift (i.e., assimilatory vs. dissimilatory L1 drift)?

4 Methods

4.1 Participants

The analysis includes two groups of L3 Spanish speakers (Instruction group and Immersion group) and a Control group. The languages were ordered based on chronology of onset of learning. All participants reported no hearing disorders and right-handedness.

The Instruction group comprised L1 Polish participants who lived in their L1 country and used their L2 (English) and L3 (Spanish) on a daily basis, but had greater exposure to the latter through instruction. The group included 14 females enrolled in Spanish studies at Polish universities during their 2–3 year of study. Their mean age was 22.5 and their mean age of onset in Spanish was 17. Due to their studies, the participants used their L3 in a formal context for about 11 hours per week. The participants also declared using Spanish outside the university for approximately 12 hours per week. Additionally, all participants reported a comparable knowledge of English and intermediate knowledge of other foreign languages including Catalan, French, Portuguese, German and Russian.

The Immersion group consisted of L1 Polish immigrants to Spain who were immersed in the L3 environment and knew L2 English at a comparable level to the Instruction group. As they were immersed in the L3 environment, they had greater exposure to their L3. This group comprised nine females and one male, whose length of residence in Spain ranged from one to 20 years, with a mean value of 10.5 years. The mean age of the participants was 37 and the mean age of onset in Spanish was 22. The participants reported using Spanish on a daily basis mostly for communication with their partners and friends as well as for entertainment and work purposes. Their use of Spanish at the time of the experiment was claimed to be approximately 37 hours per week which is nearly twice as much as that of the Instruction group. What is more, the participants declared contact with the Polish language predominantly for the purpose of contacting their family and friends as well as for entertainment for an average of 5,2 hours per week. All participants indicated a comparable knowledge of English and intermediate to advanced knowledge of other foreign languages, such as Catalan, Russian, French, Italian, German, Arabic and Ukrainian.

The Control group was composed of L1 Polish, L2 English speakers, who reported not having had any exposure to Spanish or any other language with the close front unrounded vowel /e/ at the time of the experiment. There were 12 subjects (nine females and three males), whose mean age was 24. They reported speaking English at an intermediate level. Their contact with other foreign languages (German, Russian and French) was claimed to be very limited and to have finished after high school graduation.

4.2 Stimuli

The stimulus consisted of a vowel continuum that was created by means of source-filter resynthesis performed in Praat 6.1 (Boersma & Weenink, 2019) following the Praat manual. The procedure was carried out on the /ɛ/ vowel produced in isolation by a female Polish speaker who reported hardly any knowledge of foreign languages. The mean F1 and F2 values of the obtained vowel were 685 Hz and 1790 Hz, respectively. The continuum ranged from F1 = 685 Hz to F1 = 460 Hz and consisted of six vowels, each varying from the other by 45 Hz (685 Hz–640 Hz–595 Hz–550 Hz–505 Hz–460 Hz). The files were sampled at 11,000 Hz and the peak was set to 0.14 Pa to resemble that of the original vowel. About 20 ms fragments of the obtained vowels were cut due to the disturbance to the final portion of the signal that appeared after the synthesis. The duration of all of the sounds, including the original one, was prolonged with the use of Praat Vocal Toolkit (Corretge, 2019) so that each token was 220 ms long. Other formant values, pitch, as well as all other vowel characteristics remained unchanged.

4.3 Procedure

The experiment was designed with the use of PsyToolkit (Stoet, 2010, 2017). The participants received the link to an online questionnaire which consisted of two parts: a biodata survey and an AX discrimination task. The biodata survey collected data about the participant's age, sex, place of residence, education, native language, knowledge of foreign languages as defined by the Common European Framework of Reference (CEFR), phonetic training, hearing disorders, and handedness. The participants were also asked to estimate how many hours a week they used their foreign languages in formal and informal contexts. Additionally, the survey for residents of Spain (Immersion group) also included the context of use of Polish.

After filling out the survey, the participants took part in an AX discrimination task in which pairs of the vowels from the vowel continuum appeared separated by an inter-stimulus interval of 600 ms. The AX task was chosen because of its relatively low memory load, possibility of obtaining the most reliable response times, and ease of explanation, as the study was conducted remotely through the Internet. In the task, each vowel pair was presented four times in randomised order which resulted in 84 tokens in total. The participants were asked to react to the stimuli with the use of A and L keyboard keys and decide whether the two sounds in a pair were the same or different. The reaction time was set to the maximum of 4600 ms. The task was preceded by a trial run of randomly chosen eight vowel pairs. The participants were instructed to do the experiment with headphones in a quiet room. It was explicitly stated that the differences between the sounds were small and required maximum attention. The whole experiment lasted approximately 30 minutes for each participant.

4.4 Data Analysis

Data analysis revolved around 1-step pairing to answer the formulated research questions. That is to say, the focus was placed on the following five cross-boundary pairs of sounds from the 685 Hz–460 Hz vowel continuum: 685 Hz–640 Hz, 640 Hz–595 Hz, 595 Hz–550 Hz, 550 Hz–505 Hz, 505 Hz–460 Hz, which for the purpose of the analysis were labelled as Step 1, Step 2, Step 3, Step 4, and Step 5, respectively. Correctness ratings and response times were assessed only in the aforementioned pairs of tokens. Lack of response was treated as an incorrect answer. The research questions were answered by means of comparing the percentage of correct answers and response times in the 1-step pairing. L1 drift was treated as a shift in the boundary between low and high correctness in differentiating cross-boundary pairs of sounds from the vowel continuum in the Immersion and/or Instruction groups when compared to the Control group. We then computed a two-way MANOVA with Group and Step as main effects and Group*Step as an interaction effect. The main effect of Group allowed us to determine whether there were any differences in how the Immersion and Instruction groups perceived the vowel in comparison to the Control group and, in this way, to establish whether L1 drift from their L3 took place. We also included the interaction effect of Group*Step on the correctness of discrimination and response time to see whether the type of exposure leads to different degrees and types of L1 drift.

5 Results

The overall percentage of correct answers for 1-step pairing for all groups was 22.6%. High correctness in differentiating the stimuli meant that the participant was able to hear a difference between the two sounds, and thus suggested that the participant treated the two sounds as pertaining to two different categories. In contrast, low correctness indicated lack of perceived difference, which suggested that the participant placed the two sounds in one category. The lowest percentage of correct answers was observed for the Control group (20%), while the Immersion group performed the best in the task (26%). The percentage of correct answers of the Instruction group was 22.5%.

As far as correctness is concerned, the two-way MANOVA test revealed no statistically significant effect for Group ($F = 1.206$, $p = 0.3$). However, it pointed to a significant effect for Step ($F = 11.879$, $p = 0.000$) showing a difference in discrimination of cross-boundary stimulus pairs. An interaction effect of Group*Step was also found ($F = 2.318$, $p = 0.019$), indicating a difference in discrimination of cross-boundary pairs across groups. Thus, the test results revealed that the Immersion and Instruction groups differed from the Control group in their perception of the mid front vowel by stimulus pair, suggesting that (a) L1 perceptual drift from the L3 took place, and (b) the type of exposure (immersion vs. instruction) led to different degrees of correctness for perceiving the difference at each step. Figure 1 shows the comparison of correct answers across the three groups with the focus on individual steps. Discrimination accuracy for the Immersion group was at its best in the first step (685 Hz–640 Hz) and then dwindled considerably with the decrease of the F1 value of the vowel. The Instruction group was the most accurate in the first two steps (685 Hz–640 Hz and 640 Hz–595 Hz), and its accuracy decreased as the F1 decreased. The Control group shared a similar pattern as the Instruction group but the number of correct answers in the second step (640 Hz–595 Hz) was higher and in the last three steps considerably lower. Hence, the boundary between low

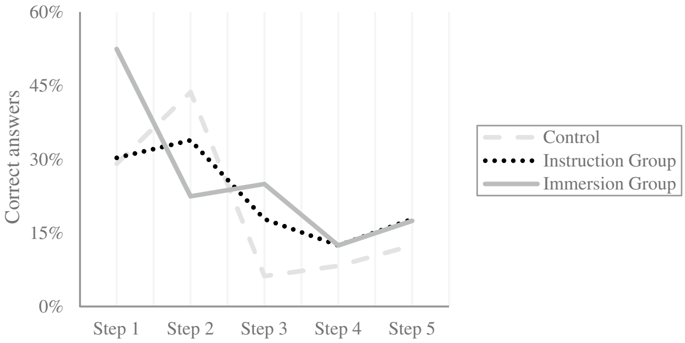

Fig. 1 Percentage of correct answers across three groups for cross-boundary pairs

correctness and high correctness for the Control group was visible between Steps 2 and 3. This suggests that the Control group places the boundary for the Polish vowel between 640 and 595 Hz in terms of the first formant. The boundary for the Polish close-mid vowel was perceived at a similar frequency by the participants of the Instruction group though the boundary was not as sharp as for the controls and extended in the direction of Step 4. The slight difference did not reach statistical significance. However, the Immersion group placed the boundary one step earlier—that is, between Steps 1 and 2, which corresponds to 685 Hz–640 Hz. Based on these results, it can be concluded that the immersion setting triggered a more substantial L1 perceptual drift than the instruction setting, and that the L1 drift was towards a greater differentiation between the L1 and L3 category—that is, it was dissimilatory L1 drift.

As for the response time (RT), the mean value for all groups in 1-step pairing was 1111.6 ms. The Control group exhibited the shortest response time (1051.5 ms), while the Instruction group took the longest to make a decision (1161.6 ms). The mean RT of the Immersion group was 1113.7 ms. The two-way MANOVA test revealed that there was no significant effect for any of the analysed factors (Group, Step, and Group*Step). Figure 2 illustrates the comparison of mean RTs across the three groups with the focus on individual steps. The mean RT of the Immersion group seems to have been the steadiest, being the longest in Step 2 and the shortest in Step 5. The Instruction group took considerably longer to answer in Step 3, while their mean RT was the shortest in Step 5. As for the Control group, their mean RT was the shortest in Step 3 and the longest in Step 4. Thus, the only difference between the groups pertains to the area between Steps 2 and 3. This difference was only observable for the Control and Instruction groups but it did not reach statistical significance. The Immersion group had a stable mean response time that was not affected by the frequency of the first formant. As this part did not yield statistically significant results, it will not be included in the analysis.

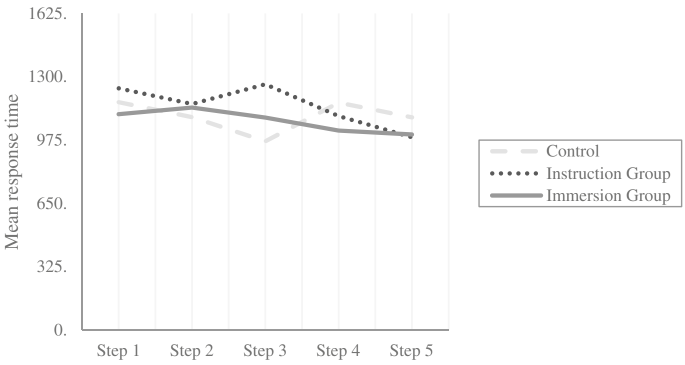

Fig. 2 Mean response time values across three groups for cross-boundary pairs

6 Discussion

The results show evidence of perceptual L1 drift from the L3 as a result of which the Polish mid front vowel /ɛ/ was perceived differently by participants with exposure to L3 Spanish than those without the exposure. This gives further confirmation for the existence of L1 perceptual drift as found by Tice and Woodley (2012), Dmitrieva (2010, 2019) and Namjoshi et al. (2015). The novel element of the current study is the source of the drift attributable to the L3. Previous attempts at capturing L3 –> L1 influence mostly showed the phenomenon as less likely to occur than L3 –> L2 influence and were limited to production in which proficiency in the L3 was lower than in the L2 (Cabrelli Amaro, 2016; Sypiańska, 2017). In the current study, capturing this direction of cross-linguistic influence has been accomplished by means of selecting a group of speakers whose L3 was the only language in their repertoire with a different mid front vowel. The drift, however, was mostly observable in the Immersion group. The Instruction group—that is, the participants who resided in their L1 country and learned L3 Spanish by means of formal instruction—, showed very small differences in their perception of the L1 Polish mid front vowel when compared to the Control group. Although these differences did not reach statistical significance, they allow to hypothesise how L1 drift could have operated in this group if the magnitude of the influence had been greater. Unlike in the Control group, the boundary between low and high correctness among the participants from the Instruction group extended towards Step 4, which is an indication of a shift in the perception of the L1 Polish mid front vowel towards a lower first formant and, thus, a higher vowel. This is a tendency towards assimilatory drift. However, in order to treat this small difference as an indication of L1 perceptual drift, a greater magnitude of influence is needed and could be obtained with a group whose exposure to L3 Spanish is more significant while remaining in the L1 country. The lack of a statistically significant L1 perceptual drift in an L1-dominant setting in the current study does not confirm previous findings for production (Herd et al., 2015) and perception (Dmitrieva, 2010). However, we believe this finding affords sufficient grounds for investigating the issue further with a group of participants residing in the L1 country but with more hours of instruction in L3 Spanish and/or lower proficiency in the L2 than in the L3.

In contrast, the Immersion group showed a statistically significant difference when compared to the other two groups (Control and Instruction groups). Thus, one major conclusion of the study is that context of learning as a contributor to greater exposure to the Lx conditions L1 perceptual drift both quantitatively and qualitatively. Firstly, the amount of L1 perceptual drift in the Immersion group was greater than in the Instruction group which confirms previous findings (Namjoshi et al., 2015). Secondly, it was qualitatively different with a dissimilatory tendency away from the L3 vowel which was the source of influence. The Immersion group showed a shift in the boundary for the L1 Polish mid front vowel when compared to the other two groups. The shift was in the direction of a higher first formant which indicates that the participants in the Immersion group perceived the Polish mid front /ɛ/ as a lower vowel in comparison to the Control and the Instruction groups.

As opposed to most findings in L1 drift studies which show cases of assimilatory drift, the current findings offer evidence of the opposite tendency. Although most data on L1 drift consist in assimilatory tendencies, in the current study the group with more exposure to the Lx was found to exhibit dissimilatory drift—that is, a tendency for greater differentiation between the L1 and the Lx property. Instead of low correctness for tokens with lower F1 as would be the case of an assimilatory tendency, the participants exhibited low correctness for the tokens with higher F1 when compared to the other two groups. However, the differentiation into assimilatory and dissimilatory drift may be property-specific as is suggested in Schuhmann and Huffman (2017). Another highly plausible explanation is based on the assumptions of the Speech Learning Model (Flege, 1995). The model predicts that "sounds in the L1 and L2 are related perceptually to one another" (p. 239) and, although the process of equivalence classification is able to block category formation for an L2 sound, with greater L2 experience, the learner may notice differences between the L1 and L2 sounds and form separate categories for them. Once separate categories are established, the learner will strive to maintain contrast between the two categories as they remain in a shared phonological space. In the first stage, under the influence of equivalence classification, the L1 and L2 sounds operate as one diaphone that is a single phonetic category used to process perceptually linked L1 and L2 sounds. The result is that the L1 and L2 sounds become more similar to each other which should evoke assimilatory L1 drift. With greater experience, as the learner notices the differences between the two sounds, a need for differentiation arises and dissimilatory L1 drift may be expected. Our study also provides a tentative suggestion of the first stage in category acquisition according to Flege as we see an assimilatory tendency in the Instruction group. Although statistically insignificant, it may point to the fact that the group is still under the influence of equivalence classification and uses (perceives) the vowels in question as a merged Polish-Spanish diaphone.

7 Implications

The results of the study have implications for understanding the process of foreign language learning and for informing second and multilingual language acquisition research and theory. Firstly, the phenomenon of changes to a native language in adults—that is, the essence of L1 drift—has been confirmed to operate at the perceptual level. This is evidence of the malleability of perception in an adult L1 that should be taken into consideration while looking at the process of learning a foreign language. A crucial contribution of the current study and a vital implication for understanding the mechanisms of foreign language learning is evidence of the L3 as a source of L1 drift. This implies that cross-linguistic influence may indeed take place among all the languages in a linguistic repertoire given sufficient exposure to the language. Although previous research pointed to a smaller likelihood of an L3 as a source language that could influence an Lx and particularly the L1, the results of our study show that it is possible in particular conditions, such as when L2 and L3 proficiency are at a similar level but the exposure to L3 is greater.

Secondly, the study adds further evidence to the understanding of the linguistic repertoire of multilinguals as highly speaker-dependent. It has been demonstrated that the phonetic properties of a language, in this case the L1, depend on the linguistic repertoire the language is part of. Each person may speak a different set of languages with different degrees of proficiency and exposure that will result in different quantities and qualities of cross-linguistic influences manifested in the L1. This has important implications for foreign language acquisition research as even greater care has to be paid to participant selection and analyses of data.

Finally, the study also shows the way that phonetic categories interact in a common phonological space depending on the amount of exposure. In particular, it depicts the consequences of the interaction for the categories of L1 sounds. These findings not only reinforce the assumptions of models of second language phonological acquisition—in particular, the Speech Learning Model—, but also provide new insights that may be the basis of models of multilingual phonological acquisition in the future.

8 Conclusions

The aim of the study was to investigate L1 perceptual drift in late-onset multilingual learners. In particular, the study sought to analyse the perception of the Polish mid front unrounded vowel /ɛ/ by L1 Polish, L2 English and L3 Spanish speakers with a comparable proficiency in their L2 and L3 but with a different type and amount of exposure to the latter. The results provide evidence of an influence from the L3 on the L1. Its source was isolated by means of looking at the mid front vowel /ɛ/, which is high in Spanish but low in the other two languages in the linguistic repertoire of the speakers. Group differences allowed to draw the following conclusions on the nature of L1 perceptual drift. Firstly, L1 drift is also manifested at the level of perception. Secondly, the amount of L1 perceptual drift depends on the context of learning which stands for exposure to the Lx. An Immersion setting affords greater Lx exposure than a formal instruction setting, thus triggering a more substantial L1 perceptual drift. Finally, contrary to most studies on L1 drift, the current analyses provide evidence of dissimilatory L1 drift away from the sound category in the source Lx. This type of drift was observed in the Immersion group and may be a consequence of greater experience in the Lx which, according to the assumptions of the Speech Learning Model, over time allows to notice the differences between the L1 and Lx sound and create a new Lx category with a tendency to maintain distance between the two. Overall, the study adds to the growing body of research on L1 phonetic drift in multilinguals focusing on speech perception—a rather unexplored area—and informs second language phonological acquisition models and models of multilingual phonological acquisition.

References

Boersma, P., & Chládková, K. (2010). Detecting categorical perception in continuous discrimination data. *Proc. Interspeech 2010*, 102–105. https://doi.org/10.21437/Interspeech.2010-60

Boersma, P., & Weenink, D. (2019). *Praat: Doing phonetics by computer* (Version 6.1) [Computer software]. http://www.praat.org/

Cabrelli Amaro, J. (2016). Testing the phonological permeability hypothesis: L3 phonological effects on L1 versus L2 systems. *International Journal of Bilingualism, 21*(6), 698–717. https://doi.org/10.1177/1367006916637287

Chang, C. B. (2010). *First language phonetic drift during second language acquisition* (ED527795) [Doctoral dissertation, University of California, Berkeley]. Dissertations/Theses—Doctoral Dissertations.

Chang, C. B. (2011). Systemic drift of L1 vowels in novice L2 learners. In W. S. Lee & E. Zee (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 428–431). City University of Hong Kong.

Chang, C. B. (2012). Rapid and multifaceted effects of second-language learning on first-language speech production. *Journal of Phonetics, 40*(2), 249–268. https://doi.org/10.1016/j.wocn.2011.10.00

Chang, C. B. (2019). Phonetic drift. In M. S. Schmid & B. Köpke (Eds.), *The Oxford handbook of language attrition* (pp. 191–203). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198793595.013.16

Corretge, R. (2019). *Praat vocal toolkit*. http://www.praatvocaltoolkit.com

de Leeuw, E., Mennen, I., & Scobbie, J. M. (2011). Singing a different tune in your native language: First language attrition of prosody. *International Journal of Bilingualism, 16*(1), 101–116. https://doi.org/10.1177/1367006911405576

de Leeuw, E., Tusha, E., & Schmid, M. S. (2017). Individual phonological attrition in Albanian-English late bilinguals. *Bilingualism: Language and Cognition*, *21*(2), 278–295. https://doi.org/10.1017/S1366728917000025

Dmitrieva, O. (2010). Phonological neutralization by native and non-native speakers: The case of Russian final devoicing. *Journal of Phonetics, 38*(3), 483–492. https://doi.org/10.1016/j.wocn.2010.06.001

Dmitrieva, O. (2019). Transferring perceptual cue-weighting from second language into first language: Cues to voicing in Russian speakers of English. *Journal of Phonetics, 73*(4), 128–143. https://doi.org/10.1016/j.wocn.2018.12.008

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. 233–277). York Press.

Flege, J. E. (2007). Language contact in bilingualism: Phonetic system interactions. In J. Cole & J. I. Hualde (Eds.), *Laboratory phonology* (Vol. 9, pp. 353–381). Mouton de Gruyter.

Flege, J. E., & Eefting, W. (1987). Cross-language switching in stop consonant perception and production by Dutch speakers of English. *Speech Communication, 6*(3), 185–202. https://doi.org/10.1016/0167-6393(87)90025-2

Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. *Language and Speech, 5*(4), 171–189. https://doi.org/10.1177/002383096200500401

Gerrits, E., & Schouten, M. E. H. (2004). Categorical perception depends on the discrimination task. *Perception and Psychophysics, 66*(3), 363–376. https://doi.org/10.3758/BF03194885

Herd, W., Walden, R., Knight, W., & Alexander, S. (2015). Phonetic drift in a first language dominant environment. *The Journal of the Acoustical Society of America, 137*(4), 2384–2384. https://doi.org/10.1121/1.4920673

Lang, B., & Davidson, L. (2017). Effects of exposure and vowel space distribution on phonetic drift: Evidence from American English learners of French. *Language and Speech, 62*(1), 30–60. https://doi.org/10.1177/0023830917737111

Liberman, A. M., Harris, K., Hoffman, H. S., & Griffith, B. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54*(5), 358–368. https://doi.org/10.1037/h0044417

Major, R. C. (1992). Losing English as a first language. *The Modern Language Journal, 76*(2), 190–208. https://doi.org/10.1111/j.1540-4781.1992.tb01100.x

Namjoshi, J., Tremblay, A., Spinelli, E., Broersma, M., Martínez-García, M. T., Connell, K., Cho, T., & Kim, S. (2015). Speech segmentation is adaptive even in adulthood: Role of the linguistic environment. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences* (paper number 0676). University of Glasgow.

Pisoni, D. B., & Lazarus, J. H. (1974). Categorical and noncategorical modes of speech perception along the voicing continuum. *Journal of Acoustical Society of America, 55*(2), 328–333. https://doi.org/10.1121/1.1914506

Sancier, M. L., & Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics, 25*(4), 421–436. https://doi.org/10.1006/jpho.1997.0051

Schuhmann, K. S., & Huffman, M. K. (2015). L1 drift and L2 category formation in second language learning. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences* (paper number 0850). University of Glasgow.

Schuhmann, K. S., & Huffman, M. K. (2017). Assimilatory and dissimilatory L1 English vowel drift in early learners of Japanese. *The Journal of the Acoustical Society of America, 141*(5), 3517. https://doi.org/10.1121/1.4987393

Schwartz, G., Dzierla, J., & Wojtkowiak, E. (2020). Laryngeal phonology and asymmetrical cross-language phonetic influence. In M. Wrembel, A. Kiełkiewicz-Janowiak, & P. Gąsiorowski (Eds.), *Approaches to the study of sound structure and speech: Interdisciplinary work in honour of Katarzyna Dziubalska-Kołaczyk* (pp. 316–325). Routledge. https://doi.org/10.4324/9780429321757-23

Schwartz, G., & Wojtkowiak, E. (2017, November 30–December 2). *Asymmetries in L2-induced phonetic drift in L1 Polish* [Conference presentation]. 11th International Conference on Native and Non-Native Accents of English, Łódź, Poland.

Stevens, K. N., Liberman, A. M., Studdert-Kennedy, M., & Öhman, S. E. (1969). Crosslanguage study of vowel perception. *Language and Speech, 12*(6), 1–23. https://doi.org/10.1177/002383096901200101

Stoet, G. (2010). PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods, 42*(4), 1096–1104.

Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology, 44*(1), 24–31. https://doi.org/10.1177/0098628316677643

Sypiańska, J. (2016). L1 vowels of multilinguals: The applicability of SLM in multilingualism. *Research in Language, 14*(1), 79–94. https://doi.org/10.1515/rela-2016-0003

Sypiańska, J. (2017). *Cross-linguistic influence in bilinguals and multilinguals.* Wydawnictwo Naukowe UAM.

Sypiańska, J. (2021). Production of voice onset time (VOT) by senior Polish learners of English. *Open Linguistics, 7*(1), 316–330. https://doi.org/10.1515/opli-2021-0016

Tice, M., & Woodley, M. (2012). Paguettes & bastries: Novice French learners show shifts in native phoneme boundaries. *UC Berkeley Phonology Lab Annual Report, 8*(8), 72–75. https://doi.org/10.5070/P79h18t4rz

Zhang, H., Chen, F., Yan, N., Wang, L., Shi, F., & Ng, M. L. (2016). The influence of language experience on the categorical perception of vowels: Evidence from Mandarin and Korean. In *Proceedings of Interspeech 2016* (pp. 873–877). ISCA. https://doi.org/10.21437/Interspeech.2016-887

Zhang, H., & Shi, F. (2014). The perceptual study of /a/ and /u/ in Beijing Mandarin. *Experimental Linguistics, 3*(1), 61–67.

Jolanta Sypiańska is Assistant Professor at the University of Szczecin, Poland. She defended her Ph.D. thesis at Adam Mickiewicz University in Poznań, Poland. She is the author of Crosslinguistic Influence in Bilinguals and Multilinguals (2017) as well as a number of articles and book chapters on multilingual acquisition of phonetics, crosslinguistic influence, and L1 drift.

Zuzanna Cal is a Ph.D. student in linguistics at Adam Mickiewicz University in Poznań, Poland. Her research interests lie in phonetic aspects of cross-linguistic influence and multilingualism as well as L1 phonetic drift. Despite the early stage of her academic career, she has conducted and co-authored a few studies pertaining to cross-linguistic influence in multilinguals.