PROCEEDINGS OF THE 10TH ANNUAL PRONUNCIATION IN SECOND LANGUAGE LEARNING AND TEACHING CONFERENCE

(ISSN 2380-9566)

EDITORS

John Levis
Charles Nagle
Erin Todey

2019, Iowa State University

TABLE OF CONTENTS

PROCEEDINGS OF THE 10TH ANNUAL PRONUNCIATION IN SECOND LANGUAGE LEARNING AND TEACHING CONFERENCE

POSTERS

TEACHING TIPS

TECHNOLOGY REVIEWS

PSLLT 2018 Conference Program

10th Annual Pronunciation in Second Language Learning and Teaching Conference

Perception and Pronunciation

September 6-8, 2018
Iowa State University
Ames, Iowa

Thursday, September 6th
(Pre-conference workshops & reception)

12:00-6:00 PM	**outside Campanile** Registration		
1:00-2:15 PM		**Oak** Workshop 1 *Investigating L2 Fluency*, Pekka Lintunen & Pauliina Peltonen	**Pioneer** Workshop 2 *Multidimensional Scaling (MDS) for Analyzing Perception Data*, Ryan Lidster
2:15-2:30 PM	**Campanile** Break		
2:30-3:45 PM		**Oak** Workshop 3 *Investigating lexical stress assignment*, Mary O'Brien	**Pioneer** Workshop 4 *Running experiments in a web browser using jsPsych*, Franziska Kruger & Danielle Daidone
3:45-4:00 PM	**Campanile** Break		
4:00-5:15 PM		**Oak** Workshop 5 *Research methods in investigating voice onset time (VOT) in second language pronunciation*, Tetsuo Harada	**Pioneer** Workshop 6 *Using surveys in pronunciation research*, Jennifer Foote
5:30-7:30 PM	**Campanile** Reception		

Friday, September 7th

Time	Campanile	Cardinal	Gold	Oak	2256 MCC
8:00 AM-4:00 PM	**outside Campanile** Registration				
8:45-9:00 AM	**Campanile** *Welcome*, John Levis				
9:00-10:00 AM	**Campanile** Plenary *High variability training in the lab and in the language classroom*, Ann Bradlow				
10:00-10:30 AM	**Campanile** *Break*				
	Campanile	**Cardinal**	**Gold**	**Oak**	**2256 MCC**
			Session 1 Presentations		
10:30-11:00 AM	*Vowel epenthesis in Korean English learners' pronunciation: at the crossroads of perception, mental lexicon, and cognitive abilities*, Hanyong Park, Isabelle Darcy	*Effects of perceptual training on vowel perception and production and implications for L2 pronunciation teaching*, Juli Cebrian, Angelica Carlet, Núria Gavaldà, Celia Gorba	*Pronunciation teaching: Whose domain is it anyways?*, Ron Thomson, Jennifer Foote	*Asymmetrical Cognitive Load Imposed by Processing Native and Nonnative Speech*, Di Liu, Marnie Reed	*Relevance of Speech Features in Building and Evaluating Automated Scoring Models*, Ziwei Zhou
11:00-11:30 AM	*Learner views on the efficiency of perceptual activities: Insights from a classroom-based study*, Anastazija Kirkova-Naskova	*The effects of task repetition on the use of epistemic stance markers: Corpus-based study*, Taichi Yamashita	*A Strategy-Based Pronunciation Model for Improving English Linking*, Veronica Sardegna	*English- and Japanese-dominant children's voice onset time (VOT) in a two-way immersion program*, Tetsuo Harada & Asako Hayashi Takakura	*"Seeing What People Hear You As": French Learners Experiencing Intelligibility Through Automatic Speech Recognition*, Aurore Mroz

11:30 AM-12:00 PM	*Learners' perceptions of a non-standard American English dialect*, Mari Sakai	*Teaching Segmentals vs. Suprasegmentals: Different Effects of Explicit Instruction on Comprehensibility*, Joshua Gordon, Isabelle Darcy	*Emergence of L2 perception: Designing and describing a high variability phonetic training study from a complex systems perspective*, Shannon Becker	*How does a speaker's face and accent affect speech processing?*, Noortje de Weers	*ASR Dictation Program Accuracy: Have Current Programs Improved?*, Shannon McCrocklin, Abdulsamad Humaidan, Idée Edalatishams
12:00-12:30 PM	*L2 French vowel production: the relationship with speech perception and phonological memory*, Solène Inceoglu	*Effects of Self-evaluation on ESL Learners' Oral Performance*, Okim Kang, Mark McAndrews	*The effect of L2 English orthographic representations on L1 Tera speakers' production and perception*, Rebecca Musa	Acquisition of prominence and tone units in English by native Japanese speakers of English: A quasi-longitudinal study, Shigehito Menjo	*Golden Speaker: Learner Experience with Computer-assisted Pronunciation Practice*, John Levis, Ricardo Gutierrez-Osuna, Evgeny Chukharev-Hudilainen, Sinem Sonsaat, Alif Silpachai, Ivana Lučić
12:30-2:30 PM	**Pioneer** Lunch (provided) & Poster Session [see list of Posters below]				
	Campanile	**Cardinal**	**Gold** Session 2 Presentations	**Oak**	**2256 MCC**
2:30-3:00 PM	*The Effect of Individual Differences on L2 Instrumental and Listener-Perceived Pronunciation*, Alyssa Kermad, Okim Kang		*Testing the malleability of teachers' judgments*, Mary O'Brien, Allison Bajt, Pavel Trofimovich, Kym Taylor Reid	*Exploring the relationship between perception and production of L2 English vowels*, Shinsook Lee, Mi-Hui Cho	*Politeness in student-professor interactions: A comparative study on the prosodic features of NS and NNS students*, Meichan Huang, Dongmei Cheng

3:00-3:30 PM	*Is perception enough? Individual differences in L2 perceptual learning and their relationship to L2 production*, Charles Nagle	*Effects of perceptual phonetic training on the perception of Korean codas by native Mandarin listeners*, Na-Young Ryu, Yoonjung Kang	*Bringing the Applied Alive in an online MA TESOL Pronunciation Course*, Betsy Parrish, Suzanne McCurdy	*Self-evaluations, perception, and production in second semester L2 French learners*, Camille Meritan	*Prosody and discourse function*, Rania Mohammed
3:30-4:00 PM	*Native listeners' evaluations on pleasantness, foreign accent, comprehensibility, and fluency toward accented talkers*, Jieun Lee, Dong Jin Kim, Hanyong Park	*Perceptual training in a classroom-setting: Phonemic category formation by Japanese EFL learners*, Ruri Ueda, Ken-ichi Hashimoto	*The Investigation of a Common Modern Spoken Arabic*, Romy Ghanem, Khaled Alharbi, Talal Alharbi	*The Perception and Production of English Initial sC(C) Clusters by Saudi ESL Learners*, Amjad Alhemaid	*Prosodic patterns in English Read by Japanese Phonetic Corpus: An Interim Report*, Takehiko Makino
4:00-4:30 PM	*A System for Analyzing and Evaluating Computer-Assisted Pronunciation Teaching Software, Websites, and Mobile Apps*, Lynn Henrichsen	*Who Follows the Rules? Differential Robustness of Phonological Principles*, John Scott	*Setting priorities for Arabic language learners: A survey of pronunciation materials in Arabic textbooks*, Ghinwa Alameen	*The perception-production interface in the acquisition of palatalized consonants in L2 Russian*, Ala Simonchyk	*The effect of instruction on receptive prosodic abilities: A meta-analysis*, Mark McAndrews

	Campanile		
4:30-4:45 PM	*Break*		
	Campanile	**Cardinal**	**Gold**
			Session 3 Presentations: Invited Speakers
4:45-5:20 PM	*Are phonological updates in the L2 mental lexicon perceptually driven?*, Isabelle Darcy & Jeffrey Holliday	*The Ripples of Rhythm: Implications for Instruction*, Wayne Dickerson	*Uses and Misues of Speech Rating Data*, Murray Munro
5:25-6:00 PM	*Investigating the phonological content of learners' lexical representations for new L2 words*, Rachel Hayes-Harb, Shannon Barrios	*Utopian Goals Revisited*, Tracey Derwing	*Discourse Intonation: Where are we now?*, Lucy Pickering
7:00-9:00 PM	Dinner		

Saturday, September 8th

8:00 AM-4:00 PM	**outside Campanile**	
	Campanile	**Cardinal**
9:00-11:00 AM	Colloquium	
	Pronunciation Research in Languages Other than English, Charles Nagle, Organizer	*Towards a Protocol for a Multilingual Corpus for Pronunciation Researchers*, Amanda Huensch & Shelley Staples, Organizers
	The role of cross-language phonetic similarity in L2 consonant learning, Anabela dos Santos Rato	
	Pronunciation in the L2 French classroom: Student and teacher attitudes, Jessica Sturm	
	Perception of French learners' mistakes, Anne Violin-Wigent and Viviane Ruellot	
	Perception of Mandarin consonants: Cross-linguistic mapping and the effect of L2 experience, Xinchun Wang	
	Campanile	
11:00-11:30 AM	Break	

	Campanile	Cardinal	Gold	Oak	3512
			Session 4 Presentations		
11:30-12:00 PM	*Template Model Account of the Intelligibility of Lexical Stress: Exemplification with Arabic-Accented English*, Ettien Koffi	*Effect of training on the perception and production of intonation: A case of Korean EFL undergraduate students*, Jin Soo Choi	*L2 Japanese Pronunciation Instruction: Its effects on improving learners' pronunciation, foreign accentedness, comprehensibility, and fluency*, Tomoko Okuno	*Design and evaluation of a computational system for learner-customized high-variability training on segmental perception in words and sentences*, Manman Qian, John Levis, Evgeny Chukharev-Hudilainen	*Transforming pronunciation through community outreach: Let me tell you their story*, Frédérique Grim
12:00-12:30 PM	*The Formation of Interactional Intelligibility due to Segmental Repair among ELF Dyads*, George O'Neal	*The Effect of Discrimination Training on Japanese Listeners' Perception of the English Coda Consonants as in 'rose' and 'roads'*, Izabelle Grenon, Chris Sheppard, John Archibald	*Development of utterance fluency and cognitive fluency and their interrelationship*, Jimin Kahng	*Informal language contact through technology and its effect on learners' use of discourse markers in oral communication*, Henriette Arndt, Christina Lyrigkou	*The role of lexical cues in the adult acquisition of L2 allophonic alternants*, Shannon Barrios, Joselyn Rodriguez
12:30-2:30 PM	Lunch break				
	Campanile	**Cardinal**	**Gold**	**Oak**	**3512**
			Session 5 Presentations		
2:30-3:00 PM	*Towards a deeper, uh, understanding of, um, L2 fluency and its [750 ms silence] correlates*, Katie Comeaux, Ron Thomson	*The use of the ICF-model in the perceptive and productive assessment and instruction for second language learners*, Ilvi Blessenaar, Lizet van Ewijk	*Corrective feedback in pronunciation teaching: A Vietnamese perspective*, Loc Nguyen, Jonathan Newton	*Drama, Skits, and Choirs to Enhance Production*, Jenelle Cox, Judy James	----

3:00-3:30 PM	*The Timing Patterns of Utterances by Native American Speakers, Cantonese Speakers, and Mandarin Speakers of English*, Bingru Chen, Jette G. Hansen Edwards	*Exploring technology in the teaching/learning of pronunciation to improve students' perception and production: teaching word and sentence stress to tertiary level students*, Nadia Kebboua & Joaquín Romero	*Oral corrective feedback timing: The case of an Iranian EFL context*, Hooman Saeli	*Effects of Japanese EFL Learners' Acoustic Short-Term Memory on English Word Reproduction Skills*, Akiko Kondo	---

	Pioneer
3:30-3:45 PM	*Break*
	Pioneer
3:45-4:35 PM	*Teaching Tips Round 1* [see list of Teaching Tips below]
4:40-5:30 PM	*Teaching Tips Round 2* [see list of Teaching Tips below]
	Pioneer
5:30-5:45 PM	*Closing,* John Levis

Teaching Tips Round 1

1. *Segmental accuracy: A recommended training sequence for moving learners from accurate perception to accurate (and automatic!) production in the stream of speech*, Monica Richards, Elena Cotos
2. *A new way of using the kazoo to teach intonation*, Colleen M. Meyers
3. *Personalizing peak vowel training in stressed syllables: A sneak peek at Blue Canoe for perception and production*, Lara Wallace & Sofía Fernandez
4. *Improving speaker comprehensibility: Using sitcoms and engaging activities to develop learners' perception and production of word stress in English*, Edna Lima, Zoe Zawadzki
5. *Developing a task-based pronunciation syllabus*, Mari Sakai
6. *Smother news or the say mold story? Coaxing the Emma cross the border*, Marsha J. Chan
7. *Improving Articulatory Gestures with Selfies*, Alison McGregor

Teaching Tips Round 2

1. *The Power of Prompts: Four Prompt Points*, Jenelle Cox
2. *Improving intelligibility: Using the three-minute thesis as a prosodic model*, Heather Boldt, Margareta Larsson

3. *Enhancing Thought Group Pedagogy Through Perception and Production*, Mark Tanner
4. *Using Tasks to Develop Comprehensibility of Spoken Second Language Spanish*, Avizia Long, Lorena Alarcón, Sergio Ruiz-Perez
5. *Practicing Pronunciation through Digital Storytelling*, Mary Ritter
6. *Listening Skills Instruction: Practical Tips for Processing Aural Input*, Marnie Reed

Poster Session Participants & Poster Titles

1	Abdulkadir Adamu	Accented Vowel Shortening and Lengthening in Nigerian English
2	Agata Guskaroska	The potential of the ASR program for facilitating vowel pronunciation practice for Macedonian learners
3	Agata Guskaroska, Joshua Taylor	A Corpus-based analysis of gendered items in pop and country music from the 90s to now
4	Alif Silpachai, Evgeny Chukharev-Hudilainen, John M. Levis, Tatiana A. Klepikova, Gabi Mitchell	The Role of Speaker Identity in High Variability Phonetic Training
5	Ane Icardo Isasa	The Effect Of A Semester-Long Phonetics Course in The Production Of L2-Spanish Vowels
6	Anna Jarosz	Secondary school learner's pronunciation needs, perceptions and attitudes
7	Anne Violin-Wigent	Longitudinal study of French liaisons and the long-term effects of explicit instruction
8	Annie Bergeron	Guatemalan seasonal workers' attitudes towards L2 French: A longitudinal study
9	Atasushi Iino, Ron Thomson	Training Japanese EFL learners to perceive English /l/, /r/, and /w/ using a cloud-based, High Variability Pronunciation Training (HVPT) application
10	Benjamin Schmeiser	On Spanish Trill Production Improvement for L1 English Learners
11	Edna Lima, Zoe Zawadzki	Suprasegmentals + sitcoms = becoming more comprehensible while having fun!
12	Fatemeh Bordbarjavidi, Erik Goodale	Learning Pronunciation through Culture
13	Ivana Lučić	Influence of Educational and Linguistic Background on Rater Perception of Second Language Oral Performance
14	Jane Lorenzen	Tonal Recall: Musical ability and toneme recognition
15	John Archibald	Intelligibility and Comprehensibility in Real Time: The Neuro- and Psycholinguistics of the L2 Parser
16	Katherine Yaw, Okim Kang, Janay Crabtree	Improving ITAs' Instructional Confidence Through Structured Contact Activities with U.S. Undergraduate Students
17	Kuo-Chan Sun	A cross-linguistic study on lexical tone processing in Mandarin L1 and L2
18	Mara Haslam, Elisabeth Zetterholm	The role of consonant clusters in English as a Lingua Franca intelligibility
19	Marcin Wojciech Telidecki	Success factors and constraints determining the acquirement of intelligible pronunciation among immigrants in the United States
20	Maria Kouti	It's not all Greek to you: Using explicit phonetic instruction in the L2 Modern Greek classroom.
21	Mark Tanner, Alisha Chugg	Using Readers Theater to Bridge the Oral Skills Gap From Perception to Production
22	Marta Nowacka	L2 pronunciation of first year English Department students: progress testing

23	Matthew Yaksic	Variations in the Production of the Neutral r-colored Vowel in L1 Spanish Speakers
24	Mikhail Zaikovskii	An Acoustic Phonetic Account of VOT in Russian-Accented English
25	Monica Richards, Elena Cotos	When perception of suprasegmental meaning varies across languages, what is a teacher to do?
26	Na-Young Ryu	Effects of L1 phonotactic constraints on L2 coda perception: A case study with native English and Mandarin learners of Korean
27	Naoko Kinoshita, Chris Sheppard	The measurement of Japanese lexical rhythm as a second language
28	Paige Gibbons, Liping Ma, Ettien Koffi	An Acoustic Phonetic Account of the Confusion between [l], [n], and [ɰ] by Some Mandarin Speakers of English
29	Pekka Lintunen, Pauliina Peltonen	Short-term gains in L2 speech during an oral skills course: Examining speech rate and fluency
30	Peter Peltekov	The Effectiveness of Implicit and Explicit Instruction on L2 German Learners’ Pronunciation
31	Raja Khan, Noor Radwan, Muhmaad Shahbaz, Aional Haryati	The role of technology in vocabulary development of EFL learners
32	Robert M. Strong	Matched-Guise As Motivator: The Power of Language Attitudes
33	Romy Ghanem, Olga Sormaz, Paula Schaefer, Qiuqu Qin	English learners' perception of intonation in different question types
34	Sondoss Elnegahy	Non-Native Learner’s Speech Perception of International Teaching Assistants in North American Universities
35	Veronica Sardegna	Increasing Pre-Service Teachers' Expert Knowledge, Effectiveness, and Agency
36	Viviane Ruellot	French stereotypical accent and pronunciation development of /p/, /t/, and /k/
37	Yoshito Hirozane	Different degrees of effects of pauses on English rate perceived by English and Japanese speakers

Nagle, C., Levis, J., & Todey, E. (2019). The changing face of L2 pronunciation research and teaching. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Ames, IA, September 2018 (pp. 1-9). Ames, IA: Iowa State University.

INTRODUCTION

THE CHANGING FACE OF L2 PRONUNCIATION RESEARCH AND TEACHING

Charles Nagle, John Levis, and Erin Todey, Iowa State University

This paper discusses changes in the field of L2 pronunciation over the past decade, including research studies, resources, and changes in methodology. To do so, it revisits the history of PSLLT over its 10 years as a conference by considering the inclusion of both research and teaching and the effects of the conference on the field of L2 pronunciation. This paper also describes changes in how L2 pronunciation research is being carried out and the general categories of the papers in the Proceedings of the 10th annual PSLLT conference.

INTRODUCTION

L2 pronunciation is a rapidly growing field, with increasing numbers of books, research articles, resources, conferences, and recognition in the larger field of applied linguistics. This growth has come with changes in the ways that pronunciation is understood, in the ways that it is being addressed in the classroom, in the ways that teacher training for pronunciation is studied, in the professional resources available for researchers of L2 pronunciation, in the expansion of concerns beyond L2 English to L2 pronunciation for other languages, and in changes in how research methodologies are evolving to address new questions.

In 2009, PSLLT debuted with 65 participants from approximately 12 different countries. That they all came to central Iowa in September was a good beginning for a new conference. Although the conference has always been about L2 pronunciation research and teaching in general, the first conference was heavily focused on English, not a surprising result given that pronunciation research at the time was also dominated by ESL/EFL/ELF concerns (Celce-Murcia et al., 1996; Jenkins, 2000; Levis, 2005; Munro & Derwing, 1995). How far we've come since then, thanks to the organizers of the conference who have each put their mark on the shape of PSLLT, and to the researchers and teachers in this rapidly growing field! The 10th PSLLT conference attracted 195 participants from more than 25 countries. The numbers of L1 and L2 combinations for pronunciation learning and teaching has increased tremendously over the conference history, and the conference has become what its name suggests, a conference and proceedings about "second language pronunciation and teaching" and not a conference about one language. We increasingly see how L2 research on the learning of Chinese tones, Japanese pitch accents and length contrasts, Spanish Voice Onset Time, using stereotypes about French accent for better pronunciation, perception of Korean codas by Mandarin speakers, and a dozen other topics help solve the many puzzles about pronunciation learning and teaching, intelligibility, comprehensibility, accentedness, perception, and production. In other words, we are on the way to what we will become as a field, and the future is bright for the field and for PSLLT.

We believe that PSLLT has played an important role in the development of the field of L2 pronunciation over the past decade. Since the 1st PSLLT conference, we now have various structural changes in the field that are closely connected to the conference.

- Almost 300 proceedings papers available freely (https://apling.engl.iastate.edu/archive/)
- A growing community of researchers and teachers who know each other and collaborate on research
- A growing number of books by researchers and teachers who are regulars at the conference (e.g., Derwing & Munro, 2015; Grantham O'Brien & Fagan, 2017; Isaacs & Trofimovich, 2017; Kang & Ginther, 2016; Kang, Thomson, & Murphy, 2018; Levis, 2018; Levis & Moyer, 2014; Levis & Munro, 2017; Moyer, 2013; Murphy, 2017; Pickering, 2018; Reed & Levis, 2015)
- A dedicated journal, the *Journal of Second Language Pronunciation*, now in its fifth year (https://benjamins.com/catalog/jslp)

CHANGES IN L2 PRONUNCIATION RESEARCH

Second language (L2) pronunciation research has flourished since the inauguration of the 1st PSLLT conference in 2009. The fact that we have now reached the critical mass of studies required for research synthesis and meta-analysis across a variety of substrands (e.g., Derwing & Munro, 2015; Lee & Plonsky, 2015; Saito & Plonsky, 2019; Sakai & Moorman, 2018; Thomson & Derwing, 2014) is a sign of the growing disciplinary maturity of the field. As any field matures, so too do its methods and practices; thus, it comes as no surprise that the two most recent conferences (2017 and 2018) featured research methods workshops and colloquia focusing on multilingual corpora construction and L2s other than English. To a certain extent, these themes are a reflection of methodological and ideological transformations taking place in applied linguistics more broadly. For instance, over the past few years, we have seen a marked increase in publications on quantitative methods and standards, such as Plonsky's (2015) edited volume on the topic, Plonsky and Oswald's (2014) publication addressing effect sizes in SLA, and Porte's (2019) book on replication research in applied linguistics. Interest in transparent reporting practices has also grown. Many L2 journals now subscribe to the Open Science Framework, awarding badges for papers that include open data or materials, and last year *Language Learning* initiated registered reports, a new type of submission that promotes methodological rigor and transparency by moving peer review to the proposal stage before data has been collected (Marsden, Morgan-Short, Trofimovich, & Ellis, 2018).

Within the growing body of literature on research methods, participant sampling practices have garnered increased attention. For example, it is well known that the social and behavioral sciences oversample individuals from Western, Educated, Industrialized, Rich, and Democratic (WEIRD) backgrounds, a tendency that could compromise the external validity and generalizability of research findings (e.g., Henrich, Heine, & Norenzayan, 2010; see Plonsky, 2017, for information specific to SLA). Beyond sociodemographic characteristics, one issue that is particularly relevant to L2 research is language sampling. Many L2 pronunciation studies involve English. In and of itself, a focus on English is not problematic since researchers are oftentimes interested in characterizing developmental patterns for speakers who share the same L1 and/or are acquiring the same L2. For example, there is a large body of work on L2 English speakers living and working in an English-speaking context (e.g., Derwing & Munro, 2013; Derwing, Munro, Foote, Waugh,

& Fleming, 2014), which reflects the sizable populations from which these speakers are drawn: the estimated 60 million individuals who speak a language other than English in the US (U.S. Census Bureau, 2015) and nearly 8 million in Canada (Statistics Canada, 2016). The opposite pairing scenario is also common. Many researchers focus on L1 English speakers who have learned another language predominantly through classroom instruction. These samples represent another population with which many of us work and to which our findings should generalize, such as the approximately 12 million K–12 students (American Councils for International Education, 2017) and 1.5 million postsecondary learners studying foreign languages in the US (Goldberg, Looney, & Lusin, 2015). In addition to studies addressing pronunciation learners, research about teachers and teacher education regarding pronunciation has heavily focused on those teaching English in ESL/EFL contexts (Breitkreutz, Derwing, & Rossiter, 2001; Burgess & Spencer, 2000; Buss, 2016; Couper, 2016; Foote, Holtby, & Derwing, 2011; Henderson et al., 2015; Macdonald, 2002), but this situation is beginning to change and we are learning about the pronunciation teaching beliefs and practices of those in non-English contexts (Huensch, 2018, 2019; Nagle, Sachs, & Zárate-Sández, 2018; Shehata, 2017; Zetterholm, 2017).

Given these trends, and the fact that this year's PSLLT included a colloquium on pronunciation in L2s other than English, we became especially interested in language sampling practices over the lifetime of the conference. We conducted an informal analysis of the languages included in conference studies published in the proceedings of the first conference (2009) and the programs of the fourth (2012) and tenth (2018) conferences (n = 148). If a study mentioned two L1s or L2s, we separated them into unique entries for the sake of analysis, and if the study mentioned more than two L1s, we coded the L1 as "Various" (n = 20). Overall, there were 17 unique L1s and 14 unique L2s, but there were only 7 studies that did not include English as either L1 or L2 (e.g., see De Meo, Pettorino, Vitalie, Cutugno, & Origlia, 2012 for Chinese learners of Italian). Figure 1 displays data for L1s and L2s that were mentioned at least twice. As is evident, we have made progress since the first conference, but there is still work to be done since English remains dominant.

One simple suggestion is to encourage colleagues working on other L2s, especially less commonly taught L2s, to attend the conference and present and publish their work, and those of us that regularly attend the conference might consider diversifying the L1s we sample, if not the L2s. Ultimately, we must reach a critical mass of studies including a variety of language pairs if we hope to derive widely generalizable findings and to shed light on the acquisition of features and phenomena, such as lexical tone and vowel harmony, that do not occur in the most frequently studied languages. At the same time, language pairs—and the dialects sampled within those pairs—is just one of many aspects of sampling to be taken into consideration.

Looking back on the past ten years, it is fair to say that PSLLT has become one of the premiere conferences for L2 pronunciation scholars, due in no small part to its integration of theory, method, and practice. Looking forward to the next decade of the conference, it seems clear that PSLLT will remain an important venue for pronunciation scholarship and will continue to grow to include a greater number of L2s.

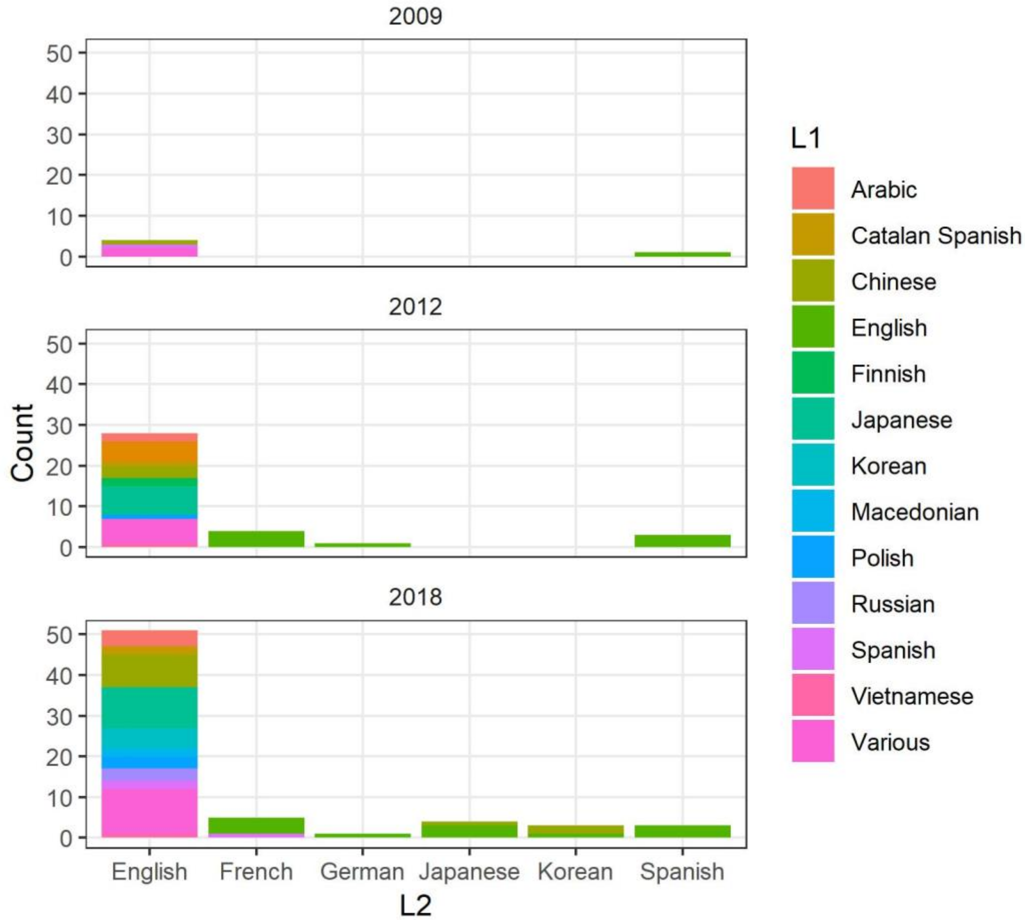


Figure 1. L1-L2 pairs in PSLLT proceedings papers (PSLLT 1) and abstracts (PSLLT 4 and 10).

THE PROCEEDINGS

This year's Proceedings, our 10th since the inauguration of PSLLT as a conference, includes nearly 50 entries. These include empirical and conceptual research, teaching tips, invited talks, research workshops, and reviews of pronunciation-related teaching technology. All but the last come from submissions to the conference proceedings.

Presentations are the largest category of papers and come from oral and poster presentations. Since the beginning, PSLLT has considered posters to be equivalent to presentations in importance, and the Proceedings often include nearly half of the papers submitted from poster presentations. Not all presenters write up their papers for the Proceedings for various reasons, including publishing their research elsewhere, having insufficient time, and feeling uncertain about how to write papers based on research presentations.

Teaching Tips, a category included since the 5th PSLLT conference, are short papers explaining ways to teach specific aspects of pronunciation for varied languages and the conceptual and/or research basis of the tips. Teaching Tips are presented in a roundtable format in which presenters stay at a round table and present their tip to a group of 8-12 participants for 7-8 minutes. At the end of the time, a bell rings, and the participants move to another table, while the presenter does their Teaching Tip for another group. This fast-paced format ends the conference, and participants walk away from the 90 minutes with 8-10 new ideas for teaching pronunciation. The Teaching Tips also allow us to keep research and teaching concerns closely related, a critical connection for L2 pronunciation (Levis, 2016a, 2016b, 2017, 2018; Levis & Wu, 2018).

The 10th Proceedings also include several Invited Talks. All previous plenary speakers and conference organizers were invited to give a talk on a topic of their choice with a longer time frame. Six presenters ultimately said yes, including two from the 1st PSLLT conference. Four of these are included in the Proceedings. The Proceedings also include Research Workshops. The pre-conference Research Workshops started in Salt Lake City in 2017. These workshops take place on Thursday afternoon before the official opening of the conference on Friday morning. They are included in the conference registration fee. Each year, they have attracted over 120 participants, a signal that they meet a growing need in the field. Two of these workshops are included in the Proceedings.

Finally, the Reviews come from work by PhD students in a course on Oral Technology and Communication at Iowa State University. They typically include reviews of little-known technology options and are included here to give them a wider readership. Reviews are often considered less important than other genres that are published, but their brevity and immediate interest to researchers and teachers suggest that they should be widely available. PhD students typically have a finger on the pulse of technology options that are currently in vogue, and their viewpoints make these a valuable addition, especially since most of the writers were involved in helping organize the conference and attended many of the sessions.

ABOUT THE AUTHORS

Charles Nagle is Assistant Professor of Spanish and Director of the Spanish Language Program at Iowa State University. His research spans three areas: how learners' pronunciation develops over time; the relationship between the perception and production of second language sounds; and, teachers' beliefs on pronunciation learning and teaching. He also published on research methods and statistics. His work can be found in venues such as *Language Learning*, *Studies in Second Language Acquisition*, and *Modern Language Journal*. At Iowa State, he teaches all levels of Spanish and upper-level Spanish linguistics courses. His email is cnagle@iastate.edu

John M. Levis is Angela B. Pavitt Professor of English at Iowa State University. He has taught and researched second language pronunciation and its teaching for many years. He is the founding editor of the *Journal of Second Language Pronunciation*, the founder of the Pronunciation in Second Language Learning and Teaching Conference, and the co-developer of http://www.pronunciationforteachers.com/. He is co-editor of several books, including the *Handbook of English Pronunciation* (Wiley), *Social Dynamics in Second Language Accent* (Degruyter), *Critical Concepts in Linguistics: Pronunciation* (Taylor & Francis), *and Intelligibility, Oral Communication and the Teaching of Pronunciation* (Cambridge University Press). His email is jlevis@iastate.edu

Erin Todey is a Ph.D. student in the Applied Linguistics & Technology Program at Iowa State University where she studies English for Academic Purposes, discourse analysis, and corpus linguistics. Erin currently works as a writing consultant and the Peer Review Group Coordinator and Facilitator in the Center for Communication Excellence. She is also the co-editor for the *TESL-EJ* Media Reviews. Her email is eftodey@iastate.edu

REFERENCES

American Councils for International Education. (2017). *The National K-12 Foreign Language Enrollment Survey Report*. Retrieved from americancouncils.org/sites/default/files/FLE-report-June17.pdf

Breitkreutz, J., Derwing, T. M., & Rossiter, M. J. (2001). Pronunciation teaching practices in Canada. *TESL Canada Journal*, 51-61.

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

Buss, L. (2016). Beliefs and practices of Brazilian EFL teachers regarding pronunciation. *Language Teaching Research*, *20*(5), 619-637.

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

Couper, G. (2016). Teacher cognition of pronunciation teaching amongst English language teachers in Uruguay. *Journal of Second Language Pronunciation*, *2*(1), 29-55.

De Meo, A., Vitale, M., Pettorino, M., Cutugno, F., & Origlia, A. (2013). Imitation/self-imitation in computer-assisted prosody training for Chinese learners of L2 Italian. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference* (pp. 90–100). Ames, IA: Iowa State University.

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. doi:10.1111/lang.12000

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

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

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*, 1-22.

Goldberg, D., Looney, D., & Lusin, N. (2015). Enrollments in languages other than English in United States institutions of higher education, Fall 2013. Retrieved from https://apps.mla.org/pdf/2013_enrollment_survey.pdf

Heinrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*(2-3), 61–83. doi:10.1017/S0140525X0999152X

Henderson, A., Curnick, L., Frost, D., Kautzsch, A., Kirkova-Naskova, A., Levey, D., Tergujeff, E., & Waniek-Klimczak, E. (2015). The English pronunciation teaching in Europe survey: Factors inside and outside the classroom. In J. Mompean & J. Fouz-González (Eds.), *Investigating English pronunciation: Current trends and directions* (pp. 260-291). New York: Palgrave Macmillan.

Huensch, A. (2018). Pronunciation in foreign language classrooms: Instructors' training, classroom practices, and beliefs. *Language Teaching Research*, 1362168818767182.

Huensch, A. (2019). The pronunciation teaching practices of university-level graduate teaching assistants of French and Spanish introductory language courses. *Foreign Language Annals*, *52*(1), 13-31.

Isaacs, T., & Trofimovich, P. (Eds.). (2016). *Second language pronunciation assessment: Interdisciplinary perspectives*. Multilingual Matters.

Jenkins, J. (2000). The phonology of English as an international language. Oxford University Press.

Kang, O., & Ginther, A. (Eds.). (2017). *Assessment in second language pronunciation*. Taylor & Francis.

Kang, O., Thomson, R., & Murphy, J. (Eds.). (2018). *The Routledge handbook of contemporary English pronunciation*. New York, NY: Routledge.

Lee, J., Jang, J., & Plonsky, L. (2015). The effectiveness of second language pronunciation instruction: A meta-analysis. *Applied Linguistics, 36*(3), 345–366. doi:10.1093/applin/amu040

Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, *39*(3), 369-377.

Levis, J., & Moyer, A. (Eds.). (2014). *Social dynamics in second language accent*. Berlin: De Gruyter Mouton.

Levis, J. (2016a). The interaction of research and pedagogy. *Journal of Second Language Pronunciation*, *2*(1), 1-7.

Levis, J. (2016b). Research into practice: How research appears in pronunciation teaching materials. *Language Teaching, 49*(3), 1-15.

Levis, J. (2017). Evidence-based pronunciation teaching: A pedagogy for the future. *Journal of Second Language Pronunciation, 3*(1), 1-8.

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

Levis, J. & Munro, M. (Eds.). (2017). *Pronunciation: Critical concepts in linguistics* (4 volumes). London: Routledge.

Levis, J. & Wu, A. (2018). Pronunciation-research into practice and practice into research. *CATESOL Journal 30*(1), 1-12.

MacDonald, S. (2002). Pronunciation-views and practices of reluctant teachers. *Prospect, 17*(3), 3-18. Retrieved from https://www.researchonline.mq.edu.au/vital/access/services/Download/mq:35597/DS01

Marsden, E., Morgan-Short, K., Trofimovich, P., & Ellis, N. C. (2018). Introducing registered reports at language learning: Promoting transparency, replication, and a synthetic ethic in the language sciences. *Language Learning, 68*(2), 309–320. doi:10.1111/lang.12284

Moyer, A. (2013). *Foreign accent: The phenomenon of non-native speech*. Cambridge: Cambridge University Press.

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, *45*(1), 73-97.

Murphy, J. (Ed.). (2017). *Teaching the pronunciation of English: Focus on whole courses*. Ann Arbor, MI: University of Michigan Press.

Nagle, C., Sachs, R., & Zárate-Sández, G. (2018). Exploring the intersection between teachers' beliefs and research findings in pronunciation instruction. *The Modern Language Journal*, *102*(3), 512-532.

O'Brien, M. G., & Fagan, S. M. (2016). *German phonetics and phonology: Theory and practice*. New Haven, CT: Yale University Press.

Pickering, L. (2018). *Discourse intonation: A discourse-pragmatic approach to teaching the pronunciation of English*. Ann Arbor, MI: University of Michigan Press.

Plonsky, L. (Ed.) (2015). *Advancing quantitative methods in second language research*. New York: Routledge.

Plonsky, L. (2017). Quantitative research methods. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 505–521). New York: Routledge.

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning, 64*(4), 878–912. doi:10.1111/lang.12079

Reed, M., & Levis, J. (2019). *The handbook of English pronunciation*. Hoboken, NJ: John Wiley & Sons.

Saito, K., & Plonsky, L. (2019). Effects of second language pronunciation teaching revisited: A proposed measurement framework and meta-analysis. *Language Learning*. doi:10.1111/lang.12345

Sakai, M., & Moorman, C. (2018). Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Applied Psycholinguistics, 39*(1), 187–224. doi:10.1017/S0142716417000418

Shehata, A. (2017). Teaching Arabic pronunciation to non-natives: Cognition and practice. In M. O'Brien & J. Levis (Eds.), *Proceedings of the 8th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Calgary, AB, August 2016 (pp. 110-120). Ames, IA: Iowa State University.

Statistics Canada. (2016). *2016 census of population – Languages spoken at home*. Retrieved from https://www12.statcan.gc.ca/datasets/Index-eng.cfm?Temporal=2016&Theme=-1&VNAMEE=Language%20spoken%20at%20home%20%28263%29&GA=-1&S=0

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

U.S. Census Bureau (2015). *Detailed languages spoken at home and ability to speak English for the population 5 years and over for United States: 2009-2013.* Retrieved from https://www.census.gov/data/tables/2013/demo/2009-2013-lang-tables.html

Zetterholm, E. (2017, December). Swedish for immigrants. Teachers' opinions on the teaching of pronunciation. In E. Babatsouli (Ed.), *Proceedings of the International Symposium on Monolingual and Bilingual Speech* (pp. 308-312). Chania, Greece.

Darcy, I. & Holliday, J. J. (2019). Teaching an old work new tricks: Phonological updates in the L2 mental lexicon. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Ames, IA, September 2018 (pp. 10-26). Ames, IA: Iowa State University.

INVITED TALK

TEACHING AN OLD WORD NEW TRICKS: PHONOLOGICAL UPDATES IN THE L2 MENTAL LEXICON

Isabelle Darcy, Indiana University, Bloomington (USA)
Jeffrey J. Holliday, Korea University, Seoul (South Korea)

We examine the dynamic relationship between perception of challenging phonological dimensions and their lexical representation in second language learners. We ask whether learners update the phonological form of words in their mental lexicon for all words simultaneously as a result of perception improvements, or whether updates are word-specific. Taking into account the trajectory of perceptual development and word learning over time, we examine this question using a lexical decision task targeting two vowel contrasts in Korean: /o/-/ʌ/ (test) and /o/-/a/ (control), the test contrast being especially challenging for L1 Mandarin learners of Korean. Participants also completed a vowel identification task, a background questionnaire, and a word familiarity questionnaire. The results confirmed that several learners had imprecise phonolexical representations, especially for words containing the test contrast. While most learners were very accurate at identifying this contrast, those with the most confusions were also least accurate in lexical decision. We also observed a trend towards word-specific phonolexical updates: words that were learned more recently were encoded more accurately than words learned earlier. The data raise the question of which lexical representations bilinguals create for words they learn, and how pronunciation instruction can help addressing phonological issues in the bilingual mental lexicon.

INTRODUCTION

This study examines how second language (L2) learners mentally store words. Lexical encoding of difficult L2 phonemic contrasts has been shown to be challenging (e.g. Dupoux, Sebastián-Gallés, Navarrete, & Peperkamp, 2008; Ota, Hartsuiker, & Haywood, 2009), but little is known about how learners update initially inaccurate lexical representations. Over time, sustained input helps the L2 phonological system develop, and processing of phonological dimensions becomes more accurate (Flege, Bohn, & Jang, 1997). Similarly, learners' lexical representations become more accurate over time (Darcy, Dekydtspotter, Sprouse et al., 2012). Given the assumed link between perceptual ability and word form learning (Pallier, Bosch, & Sebastian-Gallés, 1997; Pallier, Colomé, & Sebastian-Gallés, 2001), it is possible that improvements in lexical encoding depend on improvements in perceptual accuracy – yet this is still an open question. In this paper, we ask whether lexical representations are updated as people learn more about the phonology of their L2.

Updates to lexical representations

As adult speakers of a language, we know tens of thousands of words. We learn words our whole life, in our first language, but also in languages learned after the first. In a second language, just as in the first, some words are acquired very early, others much later. For example, for the first author who has been learning English as a second language for about 30 years, the words “hello” and “soffit” are both part of her lexicon, but one was learned in the very first days of learning English, whereas the second one was learned 30 years later, during the summer of 2018.

For each word we know (e.g. “mouse”), we also know many things about it. We know what it means, and we know what it sounds like – that is, we have a PHONOLEXICAL REPRESENTATION of the word stored in memory. We also know how it is written, that it is a noun, that its plural is *mice*, etc.

But language users don’t just learn new words, they sometimes need to change – or update – the words that they already know. For example, many of us needed to add a new meaning to the lexical representation of “mouse” in order to refer to a computer device used to position the cursor and click anywhere on the monitor. In fact, for some of us, this meaning may be the dominant meaning, while the other meaning referring to a small animal may be more weakly or not at all activated.

As this example shows, lexical representations can be updated by adding to or expanding the meanings attached to an item. The same is true for the phonological representations that we have stored in the mental lexicon – and in particular, for the words we learn in an L2. One would expect that lexical representations can evolve to reflect improved phonological knowledge and possibly show fewer instances of L1 influence in the phonolexical representation, for example. However, we don’t know whether or how this process happens, and what factors lead to updates in the phonological representations stored in the mental lexicon.

Phonolexical representations for L2 words change over time

Research evidence showing that L2 learners improve in lexical tasks suggests that it is possible to update the phonolexical representations they created for the L2 words they learned. For example, more advanced learners outperform intermediate learners in lexical decision tasks where they have to tell words from nonwords (Darcy, Daidone, & Kojima, 2013; Darcy, Park, & Yang, 2015; Darcy & Thomas, 2019). Higher proficiency also helps them detect mispronunciations in words (Simonchyk & Darcy, 2018), or helps them experience less competition from similar words (Cook, Pandža, Lancaster, & Gor, 2016; Veivo & Järvikivi, 2013). These findings suggest that learners’ lexical representations become more precise over time, and possibly become less influenced by the L1 phonology.

At the same time, a number of studies have shown that with accruing experience of the L2, learners can develop a more accurate knowledge of its phonological system. While most learners initially often misperceive phonetic and phonological dimensions, improvements can be seen in both perception and production – at least for some phonetic/phonological dimensions (e.g. Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999; Levy & Strange, 2008; Wayland & Guion, 2003).

Interestingly, these improvements in perception or production do not seem to parallel those in lexical tasks, which suggests that having acquired a contrast in perception does not guarantee accurate lexical representations (e.g. Darcy et al., 2012; Simonchyk & Darcy, 2017). In Simonchyk and Darcy, a group of intermediate and very advanced learners of Russian took part in a perception task in which they had to distinguish palatalized from non-palatalized consonants, and in a lexical task with words and nonwords involving the palatalization contrast. Knowledge of all words in the lexical task was verified with a word familiarity questionnaire. Despite high familiarity with all words, error rates on the test condition of the lexical task were high – the minimum was 30 % error, and some participants made 80–90% errors. When comparing error rates on the perception task to those in the lexical task, no clear link emerged: even one subgroup of participants with a homogeneous error rate around 15% in the perception task displayed wide-ranging error scores on the lexical task (30-90% across both groups). This means that even the listeners who were quite accurate at distinguishing the palatal and non-palatal sounds in Russian did not always know which of the two should be stored in the lexical representation of words they were very familiar with. This kind of data brings home the point that the mental lexicon in learners can be quite abstract or separate from phonetic categorization performance. We can therefore conclude that lexical updates are not always automatic, and that even at high levels of proficiency learners don't necessarily have accurate phonolexical representations.

An additional issue regarding phonolexical representations is that few studies consider item variance on lexical tasks. In a recent study (Darcy & Thomas, 2019), an item analysis revealed substantial variability in error rates between items, indicating that the lexical representations for some words seem to get updated, while other words appear to be much harder to update. Interestingly, the variability was not easily explained by item characteristics such as familiarity, frequency, loanword status, cluster type, among others.

Even though at first glance perception and word familiarity appear to have limited explanatory power in terms of the mechanisms behind lexical updates, it is possibly because both factors have mostly been considered in a "static" manner. As a large number of studies show, perception evolves over time, it is not static. Considering perception in a dynamic way might reveal a clearer relationship with the form of lexical representations (see also Nagle, 2018, on the link between perception and production). Similarly, familiarity considered alone may have limited explanatory power. But in interaction with changes in perception, it might offer an interesting picture regarding the form of lexical representations. In a nutshell, familiar words learned when perception was at a "beginner-level" may be encoded with a lack of phonological precision. If learners do not update these initial representations, these would then be reinforced over the years. By contrast, more recent familiar words, even though they may have been encountered in fewer instances, may be encoded more precisely as a result of developments in perception.

In this study, we examine the interaction of these factors in a dynamic way, considering the trajectory of perceptual development and word learning over time in learners. Simply put, this interaction could be thought of as the timing of <u>*when*</u> each word was learned. This could explain at least some of the variability previously encountered in lexical tasks even for familiar words: it is possible to imagine that words learned early (and which have been part of the mental lexicon for a long time) may be represented in less target-like ways compared to words learned more recently, if the phonolexical representations of words indeed reflect the acquisition "stage" of the L2 phonological system. We set out to investigate this idea that updates may be influenced by the timing of when a given word was learned.

We call this possibility the AGE OF WORDS HYPOTHESIS. In a nutshell, if words that contain a difficult contrast are updated according to the timing of learning, those learned early (that is, a long time ago) might be harder to update than words learned recently. In this scenario, updates are word-specific and depend on when a word was learned; updates are therefore also learner-specific, and selectively apply first to words learned *after* a new perceptual dimension has been acquired. In that case, a word's phonolexical form would also partly reflect the learner's perceptual ability regarding the difficult contrast at the time of learning. Phonolexical representations of words learned at early stages of phonological development would then reflect this earlier version of the perceptual system, while words learned more recently would reflect the perceptual progress made since. This dynamic process, by which updates first enter the lexicon through new words, and then gradually permeate the system retroactively to update older forms, is expected to take time until all words are updated, and would effectively result in old words lagging behind recent words in terms of the accuracy of lexical representations.

A different hypothesis is that word age does not play a role for the updates. Instead, learners may update their lexicon wholesale: as they acquire a new perceptual dimension (e.g. a specific vowel contrast), all lexical representations containing this dimension are updated simultaneously (PHONOLOGICAL UPDATE HYPOTHESIS). If this hypothesis is correct, there would be no effect of age.

Research questions and predictions

The following question guided the current investigation: Are updates to phonolexical representations influenced by the timing of when a given word was learned?

We hypothesize that the timing of when a word is learned indeed matters, and that there is a difference in how accurately the phonological form of words is represented in a learner's mental lexicon based on when the words were learned.

To test this hypothesis, we first establish to what extent learners have developed perceptual mastery for a pair of difficult vowels (indexed by their ability to perceptually identify each vowel); we then use a lexical decision task to probe the form of learners' lexical representations, comparing responses for words that were learned early ("old") vs. more recently ("young"). Specifically, we selected a number of test words (both old and young words) which contain the difficult vowel pair. We then manipulated these items to create nonwords by switching one vowel by its counterpart. For instance, for the difficult vowel pair /o/ and /ʌ/, an example word containing /o/ [sogɛ] 'introduce' would become the nonword [sʌgɛ] by switching out the /o/ with the /ʌ/ vowel. In a lexical decision task, the learner would ideally respond "yes" (it is a real word of Korean) to [sogɛ], but respond "no" (is it not a real word in Korean) to [sʌgɛ]. This latter response is what we call "nonword rejection", when participants successfully reject nonwords in the task. Since learners are likely to have built initially imprecise phonolexical representations for these words due to the difficult vowels they contain (Darcy, Daidone & Kojima, 2013), we interpret the failure to reject these nonwords (that is, incorrectly accepting [sʌgɛ] as a real word – see methods) to mean that their phonolexical representations are still imprecise with respect to that vowel. Likewise, successfully rejecting these nonwords indexes the accuracy with which the corresponding words are lexically represented. We compare the nonword rejection rate for nonwords based on "young" vs. "old" words. If indeed the age of words matters for phonological updating in lexical representations, the test vowels in *young* words should be encoded more accurately than in *old* words, for learners who have acquired this vowel contrast. Thus, learners should fail to reject nonwords that are based on *old* words more often than those based on *young* words, and nonword rejection rate will be higher for young words. Notably, this effect is not expected to hold for control words and nonwords (that is, those that do not contain a difficult vowel). It may even be reversed, simply because listeners are more familiar with old words and less so with newly learned words (which therefore may be harder to either accept or reject). If phonological updates happen wholesale, no such advantage for young words is expected on the test items.

METHODS

In this study, we worked with Chinese L2 learners of Korean. To examine whether these learners have acquired the difficult vowel contrast in their L2, we used a perceptual identification task for Korean vowels. The vowel contrast of interest is the /o/-/ʌ/ contrast which is reportedly difficult for Chinese learners of Korean. To probe the form of learners' lexical representations, comparing old and young words, we used an auditory lexical decision task. Participants also completed a background questionnaire, and a word familiarity and word learning history questionnaire.

Participants

Twenty-seven native speakers of Mandarin Chinese participated (mean age = 22.5 years; SD = 1.99). They were enrolled students at a major South Korean university and were living in Korea at the time of testing (mean length of residence = 1.93 years, SD = 1.52). We did not administer a separate Korean proficiency test but all participants had spent more than 1 year studying at a Korean university, where most classes are held in Korean, and they were able to fill out an extensive background questionnaire in Korean (mean age of arrival in Korea = 20.6 years, SD = 1.65).

Materials and procedure

Vowel identification task. In the vowel identification task, participants listened to 150 CV syllables excised from running speech productions. The stimuli were presented via headphones one by one, while a screen (see Figure 1) was displayed. The task was a forced choice with the 5 vowel categories (as Korean letter symbols) given as choices. They chose the letter for the vowel they heard (the IPA symbols were not displayed; they are provided here for convenience). In Korean, grapheme-phoneme correspondences for vowels are very transparent, and all learners were familiar with the letter symbols. The CV syllables contained 30 tokens each of five different vowels, including 30 /o/ and 30 /ʌ/. We look specifically at the /o/ and /ʌ/ contrast because it is the test contrast for the lexical decision task. We counted the confusions for these vowels (out of 60, expressed in %), as well as for the whole set of vowels.

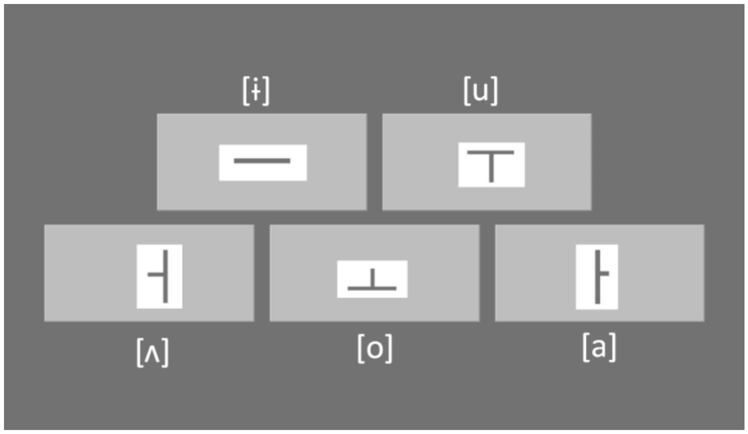


Figure 1. Screen display for the forced choice vowel identification task.

Auditory lexical decision task. This task was designed to compare the phonolexical representations for older and younger words in the same participants. Two vowel contrasts of Korean were used: /o/-/ʌ/ (test) and /o/-/a/ (control). For each contrast, stimuli were 16 words designated as “old” (i.e. likely learned a long time ago), and 16 words designated as “young” (likely learned more recently) based on Korean textbooks. Each word (8 for each vowel within the contrast) was modified to create a paired nonword by switching the vowel (e.g. /o/ for /ʌ/ and vice-versa). This resulted in 128 experimental items (64 items per contrast: 32 test, 32 control), to which 160 distractors were added. Stimuli were split into two lists, such that a word and its paired nonword never appeared in the same list. Lists were assigned randomly to participants. Each participant only heard one of the two lists, to avoid priming a response by presenting the paired item in the other list as well. The list of stimuli is presented in the Appendix A1. Since old and young items necessarily were different words, efforts were made to keep them phonologically as comparable as possible with respect to length and segments: for each old word, a young word was chosen that was as similar as possible. Table 1 displays examples of these old and young word pairs in both conditions (test and control). They were not presented as pairs in the task.

Table 1

Example stimuli for old and young words in each condition

Condition	**Old word**	(gloss)	(nonword)	**Young word**	(gloss)	(nonword)
Test	tɕ**ʌ**njʌk̚	*dinner*	tɕ**o**njʌk̚	tɕ**ʌ**nsɛ	*security deposit*	tɕ**o**nsɛ
Test	tɕ**ʌ**nhwa	*telephone*	tɕ**o**nhwa	tɕ**ʌ**npʰa	*propagation*	tɕ**o**npʰa
Test	s**o**gɛ	*introduce*	s**ʌ**gɛ	s**o**dɨk̚	*income*	s**ʌ**dɨk̚
Test	m**o**du	*all*	m**ʌ**du	m**o**sun	*contradiction*	m**ʌ**sun
Control	k**a**k*ɨm	*sometimes*	k**o**k*ɨm	k**a**nɨm	*estimate*	k**o**nɨm
Control	tɕʰ**o**rok̚	*green*	tɕʰ**a**rok̚	tɕʰ**o**biŋ	*invitation*	tɕʰ**a**biŋ

All stimuli were recorded in a sound-isolated recording booth by a phonetically trained female native speaker of Korean from Seoul. Each individual item was saved into a separate sound file for presentation by the stimulus presentation software; all files were normalized for amplitude. Participants were asked to listen to each item and decide if they heard a real word or not. They indicated their response by pressing buttons on a computer keyboard. An accurate response in this task is to say "yes" to words, and "no" to nonwords. Saying "no" to a nonword, which is always a possible word, is only possible if the phonolexical representation of the word is precise, and if the phonological difference between the word and the nonword is perceived. If – for example – a learner does not know exactly which vowel is supposed to be in the word /s**o**gɛ/ "introduce", then, s/he might think that /s**ʌ**gɛ/ (the nonword) is actually the real word. Therefore, this task allows us to probe the form of lexical representations for these words.

All items in a given list were presented auditorily through high quality headphones in a random order, only once. As soon as participants made their answer, the next item was presented after a brief delay. The task was not speeded. Prior to the test phase, participants were given 10 practice trials with feedback. Stimuli presentation was controlled by the software OpenSesame (Mathôt, Schreij, & Theeuwes, 2012). Accuracy and RT were measured, but only accuracy is used as a dependent variable in this paper.

Procedure. All procedures were approved by the Indiana University Review board. First, participants took part in the lexical decision task, followed by the vowel identification task. At the end of the experiment, each participant filled out a background questionnaire as well as a word-familiarity and learning history questionnaire, asking learners whether they knew each word, and when they thought they had learned it (Appendix A2). The entire testing session lasted about 45 mins. Participants were tested in a quiet computer room on a South Korean University campus, and were paid for participating.

RESULTS

All trials containing nonwords based on a word the listener reported not knowing (16.2% of the trials) were excluded. We also re-coded word age if, for example, a word we designated as "old" was reported by a listener to have been learned more recently. Thus, we obtained a listener-specific coding of word age, which we used in the analysis. In order to get interpretable datasets, it was necessary to exclude participants who made too many errors on the lexical decision distractors. In addition, because the coding of word-age was learner specific, some learners reported knowing too few words in certain conditions, for instance, only having two "young" items. A minimum of 5 trials in any given condition was the criterion we used for inclusion. After this exclusion procedure, the final participant sample were 13 native speakers of Mandarin who were learning Korean as L2.

Vowel identification task

Table 2 presents the rates of correct identification in the vowel identification task for each learner. Overall, all vowels were identified with relatively high accuracy, and nine out of 13 participants obtained overall average scores higher than 80% correct. When looking at the critical test contrast, we derive the accuracy in Table 2 from the number of trials on which /o/ was misperceived as /ʌ/

and vice versa. Thus, a score of 85% correct means that in 15% of the /o/ and /ʌ/ trials, the vowel was confused with the other. Participants were overall accurate in identifying these two vowels (average of 87.3% correct). Ten out of 13 participants scored 80% correct or above for this test contrast.

Table 2

Accuracy (%) in the vowel identification task (test vs. all contrasts) for each participant

Learner	27f1	25f1	18f2	22f2	13f1	05f1	24f2	09m1	11f1	10f2	07f1	14f2	21f1
/o/-/ʌ/	65	68.3	75	80	85	86.7	88.3	93.3	95	98.3	98.3	100	100
All vowels (incl. /o/-/ʌ/)	61.7	60	70	78.3	80	85	85	88.3	91.7	93.3	91.7	88.3	93.3

Auditory lexical decision task

In the current analysis, we defined “accuracy” as the nonword rejection rates in each trial type (test vs. control). Generally, we expected an effect of trial type, that is, higher nonword rejection rates for control items compared to test items. This first prediction was confirmed. A mixed effects logistic regression model with fixed effects of *word age* and *trial type* revealed a significant effect of trial type, where test items were less accurate than control items; $\beta = 1.17$, $p < .001$).

Second, we examine any potential effects of word age by comparing responses for items based on old words vs. young words (declaring the variable of word age in the regression model). For the test contrast, accuracy for nonwords based on old words was on average 13.7% lower than for those based on young words, whereas for the control contrast, this difference was much smaller: accuracy for items based on old words was 6% lower than for items based on young words. Figure 2 shows the mean accuracy in each trial type and word age: control vs. test, and old vs. young.

Descriptively, the hypothesis that word age matters for updates is supported. Statistically however, it was not. The tendency observed in the test trials is in the predicted direction: Nonwords based on young words are rejected more successfully than nonwords based on old words. This might indicate that young words are easier to update (to correct), and therefore nonwords based on these are easier to reject. However, this trend did not produce a significant interaction between age and condition, possibly due to the small sample size and the fact that variability between participants is large. The mixed effects logistic regression model revealed a significant effect of trial type on accuracy scores (test < control), but no significant effect of word age, and no significant interaction between age and trial type. Together with the correlation presented above, the absence of a statistically significant effect of word age tentatively supports the phonological update hypothesis, in which individual lexical forms appear to be updated wholesale as learners learn more about the phonological system of their L2.

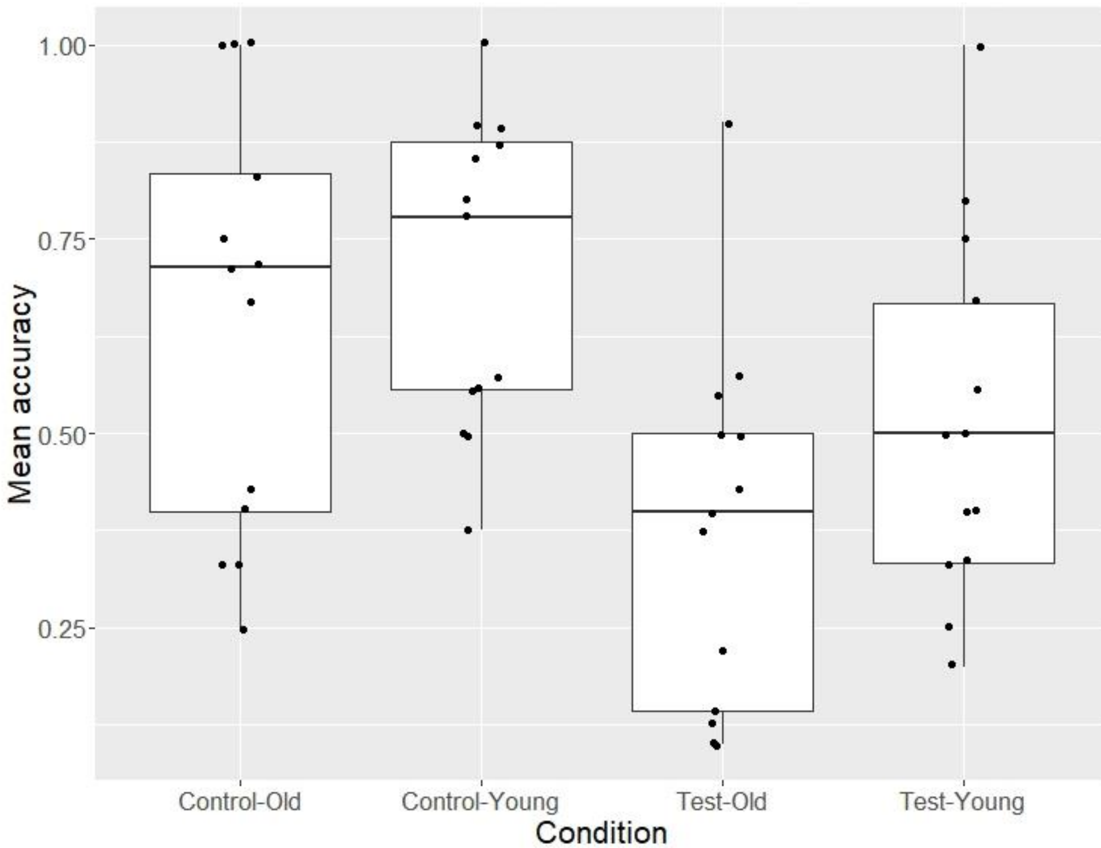


Figure 2. Boxplot/strip chart for mean accuracy (nonword rejection rate) in the four experimental conditions (trial type x word age). Each dot represents one participant (n = 13).

Finally, we examined the link between vowel identification and lexical decision accuracy on test items (across both old and young items). A two-tailed non-parametric Spearman correlation showed that vowel identification accuracy was positively correlated with accuracy in lexical decision ($r_s = .71$, n = 13, $p = .007$). The scatterplot in Figure 3 shows the relationship between the two measures. The red square across the top highlights that even at very high levels of identification accuracy for this contrast, performance on the lexical decision task is very variable (ranging from about 40% to 95% correct), thus again indicating that even near perfect identification does not guarantee lexical accuracy.

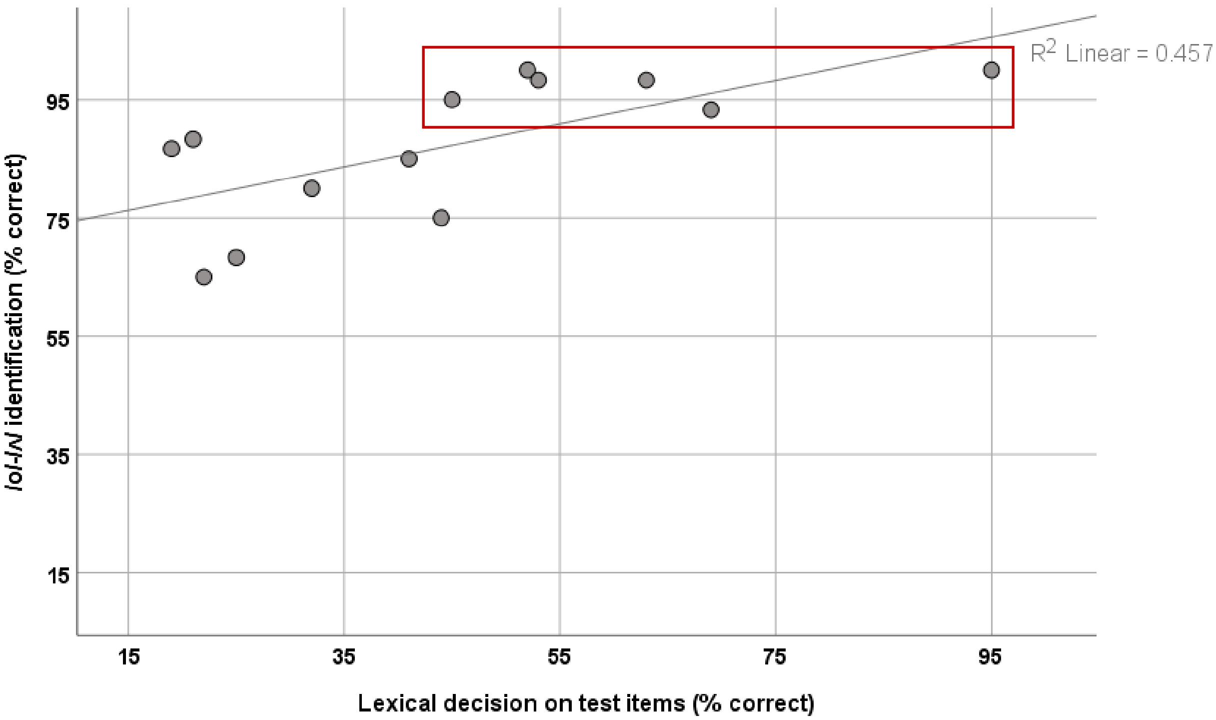


Figure 3. Scatterplot showing the relationship between identification and lexical decision scores. Each dot is one participant.

In the global regression analysis above, all participants were included. Yet, word age effects, if any, are expected to be emerging mainly for those participants who have acquired the contrast successfully, because in the word age hypothesis, any effects *depend on* perceptual mastery of a contrast. Therefore, it is important to examine this pattern in participants with an excellent perceptual mastery of the vowel contrast. Figure 4 shows the relationship between perceptual identification accuracy for the test vowels and the lexical decision scores on the test items, split by word age (young: empty vs. old items: filled). Participants are ranked on the x-axis from least to most accurate in the perceptual identification task for this contrast (%correct). We found that listeners who confused the contrast more often were also significantly more likely to incorrectly accept nonwords based on *both* old and young words (both $R^2 = 0.352$, $p < .02$), perhaps suggesting that in their mental lexicon, words do not encode the contrast reliably, regardless of word age. However, a difference between old and young test items emerges more reliably for those participants with the highest mastery of the contrast: for all participants with a perceptual identification score above 90%, young items are responded to more accurately than old ones. This relationship was not visible in the control contrast.

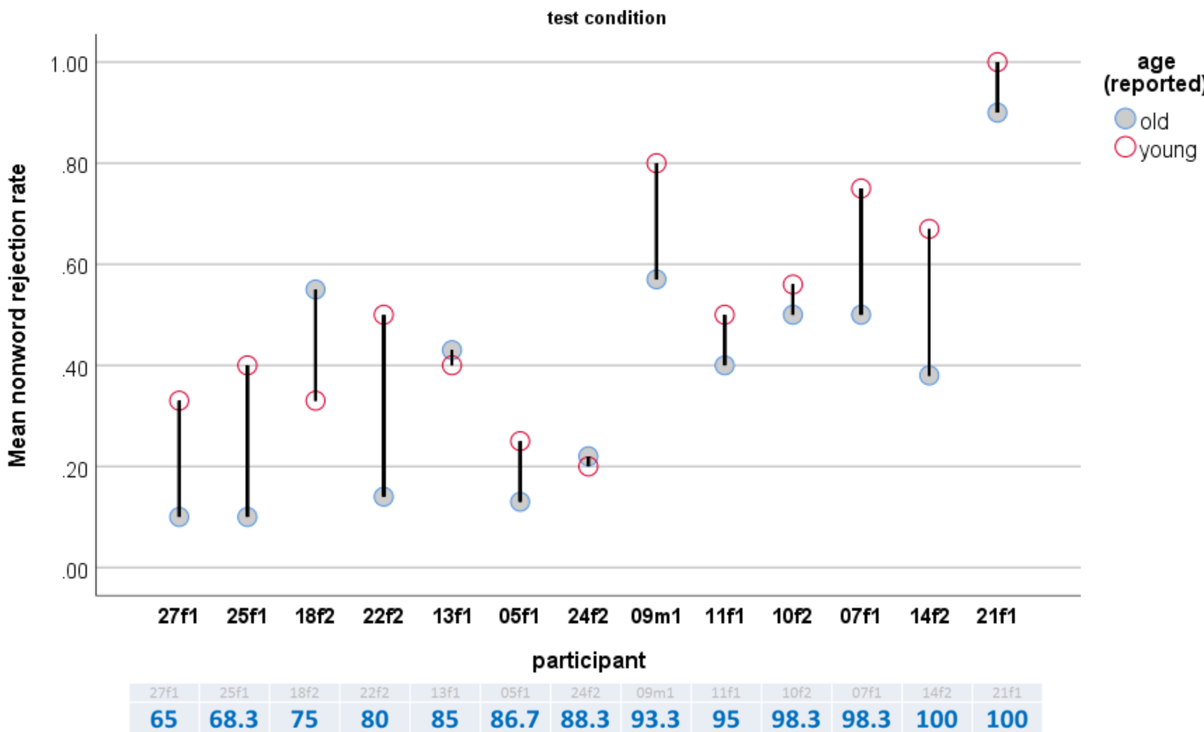


Figure 4. Relationship between perceptual identification accuracy for the test vowel and the lexical decision scores for test items, split by word age. The bottom numbers indicate individual identification accuracy (%) for the test contrast, ranked from lowest to highest.

DISCUSSION

The absence of a statistically significant effect of word age tentatively supports the phonological update hypothesis. Phono-lexical representations appear to be updated wholesale as learners learn more about the phonological system of their L2, given the fact that we observed a correlation between perception accuracy and nonword rejection rate for both young and old items. Yet, a difference by word age becomes most clearly visible in participants with the highest mastery of the vowel contrast. Thus, we cannot fully reject the possibility that an effect of age might emerge. However, at this stage this effect is small and must remain tentative.

Despite the preliminary nature of these findings, they suggest that potentially, once a contrast is acquired and part of a learner's phonological knowledge, it is first represented in the phonolexical representation of more recent words, and little by little, may reach more entrenched representations. The limited sample size of this investigation limits the conclusions we can draw, but our exploratory study raises interesting questions.

The most important question raised by our findings concerns the inventory of lexical representations in the bilingual mental lexicon. It is indeed possible that phonolexical representations are not actually updated (corrected), but rather that new ones are added to previously existing ones, and co-exist for a time. Accordingly, the earlier representations remain accessible during word recognition. One benefit of this kind of scenario is that word recognition would be facilitated for both target-like input as well as for non-target-like input (which follows the L1 phonological specifications). Concretely, this means that an L2 learner who builds a new accurate representation next to an earlier one may be as fast to recognize the word when spoken by a native speaker as when spoken by another L2 learner who confuses the same contrast in their production. In the long term however, L2 learners ideally need to inhibit the activation of – or even fully erase – the earlier phonolexical representations. Whether learners are able to do this at all remains a question for future research.

Our findings further indicate that older lexical representations may be more resistant to updates than recent lexical representations. If this finding is confirmed, it will need to be reconciled with other data showing that more entrenched lexical representations also lead to higher accuracy and less variability in specific lexical tasks (e.g. Cook et al., 2016). Our results suggest the opposite. However, the two possibilities need not be incompatible: they may derive from different tasks targeting different phonological contrasts and proficiency levels. In particular, findings such as ours are likely to apply mostly to difficult phonological dimensions, whereas Cook and colleagues obtained data across a wider range of phonological dimensions, not all of them being perceptually challenging.

Finally, our findings relate to L2 pronunciation in two main ways. The first is that globally, lexical updates appear to rely on improvements in phonological knowledge; yet, even among the learners who have mastered the nonnative vowel contrast in perception, accurate or updated lexical representations were not guaranteed. If these findings are indeed solid (similar findings were obtained in Simonchyk & Darcy, 2017), they lead us to assume a kind of ‘hierarchy’ in the order of acquisition such that phonological knowledge of sounds may precede phonolexical updates, and not the other way around (that is, in L2, phonological knowledge may not emerge from inferences over the mental lexicon, a scenario that has been proposed for L1 acquisition by, among others, Munson, Edwards and Beckman [2005, p. 198]: “Higher level phonological knowledge emerges as a consequence of word learning and serves to facilitate future word learning”). While this assumption clearly needs to be fully tested for L2 learners, given its relevance for instruction, it also begs the question of why updating lexical representations lags behind improvements in phonological knowledge. One possible answer is that the kind of perceptual improvements underlying accurate identification in a task such as ours are not linked to actual word representations as they occur, thus hindering a direct influence on these word representations. Another possible answer is that updating lexical entries takes more time or more directed effort. Thus, teachers should be aware of this potential discrepancy in learners’ productions, which are likely to reflect fuzzy lexical representations for a long time. This would not necessarily mean that learners have not yet acquired the corresponding phonological knowledge, it may just indicate that some words have not yet been updated. Clearly, the precise mechanisms by which perceptual and phonolexical improvements take place as well as their time course are yet to elucidate. But there is potentially a role for pronunciation instruction to help bridge the gap between perception and lexical representations, and this is the other way in which our findings relate to pronunciation instruction.

Our findings highlight the need for integrating phonological knowledge into lexical knowledge. Importantly, we are not suggesting that teachers should first ensure learners have acquired the phonological system of the L2 before they start learning words. This approach would not only be unrealistic, it may also be counterproductive, since in both L1 and L2, learning words might still be a driving force in developing the phonological system in the first place (Munson et al., 2005). A much more effective approach would be to incorporate pronunciation instruction about words at the earliest possible time by integrating new vocabulary teaching with pronunciation, and by revisiting known words often while focusing on their phonological form. This approach might prove successful in preventing too many fuzzy lexical representations from getting set up in the first place.

ACKNOWLEDGMENTS

The authors wish to thank Zoie Hancock and Joshua Lee for their help with the stimuli creation. We also thank audience members at the 10th PSLLT conference in Ames, IA and at LabPhon 16 in Lisbon, Portugal, for stimulating discussion.

ABOUT THE AUTHORS

Isabelle Darcy is associate professor of second language psycholinguistics in the Department of Second Language Studies at Indiana University. She obtained a PhD in Linguistics and Cognitive Science from the Ecole des Hautes Etudes en Sciences Sociales in Paris (France) and from the Johannes Gutenberg University in Mainz (Germany). Her research focuses on the acquisition of second language phonology, pronunciation instruction, and on L1/L2 speech perception and word recognition.

Indiana University
Department of Second Language Studies
Morrison Hall 231
1165 E. 3rd St.
Bloomington, IN 47405, USA
Phone: +1 812 855 0033
E-mail: idarcy@indiana.edu

Jeffrey J. Holliday is assistant professor in the Department of Korean Language and Literature at Korea University in Seoul, South Korea. He obtained a PhD in Linguistics from The Ohio State University. His research interests include laboratory phonology, perception and production of Korean as a second language, socio-phonetics, and individual differences.

Department of Korean Language & Literature
Korea University
145 Anam-ro Seongbuk-gu
Seoul 02841
South Korea
E-mail: holliday@korea.ac.kr

REFERENCES

Bradlow, A., Akahane-Yamada, R., Pisoni, D., & Tohkura, Y. i. (1999). Training Japanese listeners to identify English /r/and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics*, 61, 977-985. doi:10.3758/BF03206911

Cook, S. V., Pandža, N. B., Lancaster, A. K., & Gor, K. (2016). Fuzzy Nonnative Phonolexical Representations Lead to Fuzzy Form-to-Meaning Mappings. *Frontiers in Psychology, 7*(1345). doi:10.3389/fpsyg.2016.01345

Darcy, I., Daidone, D., & Kojima, C. (2013). Asymmetric lexical access and fuzzy lexical representations in second language learners. *The Mental Lexicon, 8*, 372-420.

Darcy, I., Dekydtspotter, L., Sprouse, R. A., Glover, J., Kaden, C., McGuire, M., & Scott, J. H. G. (2012). Direct mapping of acoustics to phonology: On the lexical encoding of front rounded vowels in L1 English-L2 French acquisition. *Second Language Research*, *28*, 5-40. doi:10.1177/0267658311423455

Darcy, I., Park, H., & Yang, C.-L. (2015). Individual differences in L2 acquisition of English phonology: The relation between cognitive abilities and phonological processing. *Learning and Individual Differences*, *40*, 63-72. doi:http://dx.doi.org/10.1016/j.lindif.2015.04.005

Darcy, I., & Thomas, T. (2019). When *blue* is a disyllabic word: Perceptual epenthesis in the mental lexicon of second language learners. *Bilingualism: Language and Cognition*, 1-19. doi: 10.1017/S1366728918001050 (first view).

Dupoux, E., Sebastián-Gallés, N., Navarrete, E., & Peperkamp, S. (2008). Persistent stress 'deafness': The case of French learners of Spanish. *Cognition, 106*, 682-706.

Flege, J. E., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, *25*, 437-470.

Levy, E. S., & Strange, W. (2008). Perception of French vowels by American English adults with and without French language experience. *Journal of Phonetics, 36*, 141-157.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314-324. doi:10.3758/s13428-011-0168-7

Munson, B., Edwards, J., & Beckman, M. E. (2005). Phonological knowledge in typical and atypical speech-sound development. *Topics in Language Disorders*, *25*, 190-206.

Nagle, C. L. (2018). Examining the temporal structure of the perception–production link in second language acquisition: A longitudinal study. *Language Learning*, *68*, 234-270. doi:10.1111/lang.12275

Ota, M., Hartsuiker, R. J., & Haywood, S. L. (2009). The KEY to the ROCK: Near-homophony in nonnative visual word recognition. *Cognition, 111*, 263-269.

Pallier, C., Bosch, L., & Sebastian-Gallés, N. (1997). A limit on behavioral plasticity in speech perception. *Cognition, 64,* B9-B17.

Pallier, C., Colomé, A., & Sebastian-Gallés, N. (2001). The influence of native-language phonology on lexical access: exemplar-based versus abstract lexical entries. *Psychological Science, 12*, 445-449.

Simonchyk, A., & Darcy, I. (2017) Lexical encoding and perception of palatalized consonants in L2 Russian. In M. O'Brien & J. Levis (Eds). *Proceedings of the 8th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Calgary, AB, August 2016 (pp. 121-132). Ames, IA: Iowa State University.

Simonchyk, A., & Darcy, I. (2018) The Effect of Orthography on the Lexical Encoding of Palatalized Consonants in L2 Russian. *Language and Speech, First Published March 18, 2018*. doi: 10.1177/0023830918761490

Veivo, O., & Järvikivi, J. (2013). Proficiency modulates early orthographic and phonological processing in L2 spoken word recognition. *Bilingualism: Language and Cognition, 16*(04), 864-883. doi:10.1017/S1366728912000600

Wayland, R., & Guion, S. (2003). Perceptual discrimination of Thai tones by naive and experienced learners of Thai. *Applied Psycholinguistics, 24*, 113-129.

APPENDIX

Table A1

List of stimuli used in the lexical decision task

	Test /o/ words					
	Old words			**Young words**		
	IPA	**(gloss)**	**(nonword)**	**IPA**	**(gloss)**	**(nonword)**
1	norɛ	*song*	nʌrɛ	nodoŋ	*labor*	nʌdoŋ
2	modu	*all*	mʌdu	mosun	*contradiction*	mʌsun
3	koŋbu	*study*	kʌŋbu	koŋdʑi	*notification*	kʌŋdʑi
4	koŋtɕ*a	*free*	kʌŋtɕ*a	koŋhak̚	*engineering*	kʌŋhak̚
5	totɕʰak̚	*arrive*	tʌtɕʰak̚	tosʌ	*book*	tʌsʌ
6	toŋsɛŋ	*younger sibling*	tʌŋsɛŋ	toŋtɕʰaŋ	*alumnus*	tʌŋtɕʰaŋ
7	sogɛ	*introduce*	sʌgɛ	sodɨk̚	*income*	sʌdɨk̚
8	onɨl	*today*	ʌnɨl	onsu	*hot water*	ʌnsu
	Test /ʌ/ words					
1	tɕʌnjʌk̚	*dinner*	tɕonjʌk̚	tɕʌnsɛ	*security deposit*	tɕonsɛ
2	mʌndʑʌ	*first*	mondʑʌ	mʌŋgɛ	*sea squirt*	moŋgɛ
3	pʌls*ʌ	*already*	pols*ʌ	pʌltɕʰik̚	*penalty*	poltɕʰik̚
4	ʌlgul	*face*	olgul	ʌlluk̚	*stain*	olluk̚
5	tɕʌnhwa	*telephone*	tɕonhwa	tɕʌnpʰa	*propagation*	tɕonpʰa
6	tɕʰʌŋso	*clean*	tɕʰoŋso	tɕʰʌŋtɕʰun	*adolescence*	tɕʰoŋtɕʰun
7	ʌnni	*older sister*	onni	ʌndʌk̚	*hill*	ondʌk̚
8	tɕʌmɕim	*lunch*	tɕomɕim	tɕʌmtɕʰa	*gradually*	tɕomtɕʰa
	Control /o/ words					
	Old words			**Young words**		
	IPA	**(gloss)**	**(nonword)**	**IPA**	**(gloss)**	**(nonword)**
1	komin	*trouble*	kamin	kotɕʰɨŋ	*high floor*	katɕʰɨŋ
2	koguŋ	*palace*	kakuŋ	kogal	*depletion*	kagal
3	mok̚tɕʌk̚	*goal*	mak̚tɕʌk̚	mok̚s*um	*life*	mak̚s*um
4	motɕip̚	*recruit*	matɕip̚	mobʌm	*model*	mabʌm
5	momsal	*flu*	mamsal	momtɕit	*gesture*	mamtɕit̚
6	tɕʰorok̚	*green*	tɕʰarok̚	tɕʰobiŋ	*invitation*	tɕʰabiŋ
7	onmom	*whole body*	anmom	onɕil	*heated room*	anɕil
8	otɕ*aŋ	*closet*	atɕ*aŋ	ogok̚	*five grains*	agok̚
	Control /a/ words					
1	kak*ɨm	*sometimes*	kok*ɨm	kanɨm	*estimate*	konɨm
2	kasɨm	*chest*	kosɨm	kamum	*drought*	komum
3	nampʰjʌn	*husband*	nompʰjʌn	namgɨk̚	*south pole*	nomgɨk̚

4	tanpʰuŋ	*maple*	tonpʰuŋ	tansʌ	*cue*	tonsʌ
5	tap̚tɕ*aŋ	*reply*	top̚tɕ*aŋ	tap̚ɕ*in	*reply (formal)*	top̚ɕ*in
6	annɛ	*information*	onnɛ	angɛ	*fog*	ongɛ
7	tɕʰaŋmun	*window*	tɕʰoŋmun	tɕʰaŋdʑo	*creation*	tɕʰoŋdʑo
8	hanɨl	*sky*	honɨl	hatɕʰɛ	*lower body*	hotɕʰɛ

Note. Darker shading are items assigned to list 1, those with no shading to list 2.

Date____________________ IRB Study #_________

Participant ______________

N° Group ______________

When did you learn these words?

Please estimate when, during your Korean learning experience, you have learned these words:

* If the word is among those you learned first → please check box #1: "In my first year"

* If you have learned this word some time ago, but not in your first year → check #2 "in-between"

* If you have learned the word not very long ago → check #3: "quite recently"

Item	In my first year	in-between	quite recently	Item	In my first year	in-between	quite recently
song				labor			
all				notification			
study				engineering			
free				book			
arrive				alumnus			
younger sibling				income			
introduce				hot water			
today				security deposit			
dinner				sea squirt			
first				penalty			

Figure A2. Sample questionnaire items used to estimate listener-specific word age.

Derwing, T. M. (2019). Utopian goals for pronunciation research revisited. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching conference*, ISSN 2380-9566, Iowa State University, September 2018 (pp. 27-35). Ames, IA: Iowa State University.

INVITED TALK

UTOPIAN GOALS FOR PRONUNCIATION RESEARCH REVISITED

Tracey M. Derwing, University of Alberta, Simon Fraser University

In 2009, I gave a presentation at the inaugural PSLLT entitled Utopian Goals for Pronunciation Teaching. Here I revisit those goals to see how far we have come. Pronunciation is no longer the Cinderella of applied linguistics research; in fact, it is the Belle of the Ball, in that not only are many more PhDs graduating with a focus on L2 pronunciation, but established academics whose primary interests are elsewhere are now collaborating with others to examine pronunciation in relation to their own research specialty. We have seen massive increases in empirical studies on L2 pronunciation, as well as the establishment of a journal devoted to L2 pronunciation issues. In addition, innovations in technology devoted to pronunciation improvement have emerged. However, there is still considerable room for improvement and development in our field. I will address the Utopian goals identified in the original paper, outlining progress thus far, and suggesting ways forward. The goals are: increased attention to pronunciation from researchers; a stronger focus on teacher education; appropriate curriculum choices; a stronger focus on intelligibility/comprehensibility; more useful software/other technology; a focus on NS listeners; no more scapegoating of accent; and better strategies for integrating newcomers into the community.

The first PSLLT was the impetus for this paper: in 2009 I was invited to give a presentation on what I thought would contribute to better learning experiences for students (Derwing, 2010). Pronunciation at that time was still the Cinderella of second language acquisition, and now it is the Belle of the Ball. Still, we have not yet reached a happy ending. Considerably more work is necessary before we get there. Here I will revisit the goals set in 2009 to examine how far we have come as a field.

Although the presentation ten years ago dealt with “Utopian goals”, the goals can also be characterized as a to-do list, because several are quite do-able.

The first goal was increased attention to pronunciation from researchers. Of course, there is always an ongoing need for research, but it is clear that the last decade has witnessed monumental growth in the number of research projects on second language (L2) pronunciation, in part because of the establishment of PSLLT and the *Journal of Second Language Pronunciation*. Two overview studies, one a meta-analysis (Lee, Jang & Plonsky, 2015) and the other a narrative review (Thomson & Derwing, 2015) appraised the burgeoning research in this area. Since these articles were published there has been no reduction in interest; each year several more PhDs graduate in our field. In 2017, the American Association of Applied Linguistics received so many submissions on pronunciation issues that in 2018 they created a new area strand entitled Phonology/Phonetics and Oral Communication. Furthermore, attendance at PSLLT has grown from 60 to 195 registrants. It is also heartening that people whose first area of interest is another aspect of second language acquisition are now including L2 pronunciation in their own research. For instance, Loewen and Isbell (2017) and Ruvivar and Collins (2018) have recently published L2 pronunciation studies. Both Shawn Loewen and Laura Collins are senior researchers whose primary focus is grammar. Clearly, the goal for more attention from researchers has been met! However, several areas remain under-researched, especially within the classroom. For instance, more focus on learners, and the individual and combined effects of different PI activities would be useful.

The second goal was a stronger focus on teacher education. In April, 2017, John Levis established a website intended for teachers, populated with accessible essays on key concepts and teaching techniques: pronunciationforteachers.com. A resources section includes links to webinars and the like. This website is still in a nascent stage but it is growing, and it is exceptional in that the evidence-based essays are all written by leading researchers and expert practitioners in the field. Moreover, in the last few years, several textbooks intended for teachers have been published, all of which are informed by research, including Derwing and Munro (2015), Grant (2014), Levis (2018), Murphy (2013), and Pickering (2018). This is by no means an exhaustive list, whereas at the time of the first PSLLT, there were very few choices of texts for teacher preparation in L2 pronunciation instruction. Teacher education in this area still lags behind, however, because of limited access to courses on 'how to teach pronunciation' in the ivory towers. Many university programs assume that a course in general linguistics or phonology is enough; it is most emphatically *not* enough! In my own experience, at a university which has regularly offered a 'teaching L2 pronunciation' course, students were required to first take a course in linguistics. Although that was useful, most preservice teachers were unable to see exactly how their linguistic knowledge base related to teaching; they needed explicit instruction regarding pronunciation instruction research and teaching to inform their own practice. I am optimistic that in the next decade more courses of this type will be developed as more PhD graduates whose primary interest is L2 pronunciation are hired in university positions. In the meantime, a census and comparison of existing language teacher preparation programs would serve to identify both gaps and good practices.

Goal #3 related to appropriate curriculum choices. Stand-alone pronunciation classes can be helpful, particularly if several students share similar difficulties. However, programs may not have sufficient numbers to run stand-alone classes, or students may need to work on other aspects of their L2 as well, which suggests that pronunciation is best integrated into general listening and speaking classes. My sense from reading many classroom-based studies is that students who are receiving pronunciation instruction are mostly registered in stand-alone classes and that we still do not see much pronunciation integrated into general language classes. Years ago, Levis and Grant (2003) pointed to the lack of systematicity in the inclusion of pronunciation in general ESL classes and provided suggestions for ways to incorporate pronunciation. More calls have come out to this effect, but there has been little progress on this front. Numerous studies exist on teachers' attitudes towards pronunciation, but it would be helpful to research and highlight successful integration of PI into general language classes.

Another curriculum choice is whether students have the opportunity for exposure to multiple voices from several L2 accents and dialects. A host of studies have shown that High Variability Phonetic Training (HVPT) is beneficial for learners' perception and sometimes production (Thomson, 2018). We now see more HVPT, mostly because of developments in technology, and a growing awareness that pronunciation is not just about what the students can produce, but also what they can perceive, and clearly they need to be able to understand a range of accents and dialects.

Yet another aspect of curriculum choices is assessment. In the USA, assessments exist for international teaching assistants (ITAs), but in many types of language programs, and certainly in most of Canada, programs and teachers tend to avoid assessing pronunciation. However, if it were tested, it would be taught. Thus, I am in favour of the development of assessment tools for pronunciation. We know there will be washback, so it is incumbent on us to design good tests. Kermad and Kang (2018) recently compared low, medium and high stakes tests of L2 pronunciation, and determined that students tend to do best on low stakes tests. They encourage teachers to vary the content and design of tests to match high and low stakes assessments to give students practice such that their productions will remain at their best. Moreover, better protocols for classroom needs assessment are in order.

Goal #4 was a focus on intelligibility and comprehensibility rather than accent. There is definitely more discussion of the first two speech dimensions than several years ago, but accent is still winning the day. Accent reduction programs have proliferated: in a quick search in 2018, I found 11,500,000 hits on Google vs. 633,000 in 2008. Many of the providers are fear-mongers whose practices will not help the clients at all, and in fact, may be detrimental (Derwing & Munro, 2015). We see improvement in a focus on comprehensibility and intelligibility among researchers, but, in a review of 75 pronunciation instruction studies (Thomson & Derwing, 2015), 63% aligned with the Nativeness Principle and 24% with the Intelligibility principle, while 13% had elements of both (See Levis, 2005 for a discussion of the two principles). And that is among applied linguists; in the last ten years, social psychologists have started to take a great interest in foreign accentedness, with little regard for even the existence of other speech dimensions. Of course, an L2 accent can have significant socio-cultural consequences, but it is not an all or nothing phenomenon, and more nuanced considerations involving intelligibility and comprehensibility are warranted.

Goal #5 was the development of more useful software and other technology. Amazing advances have been made in this area. Ron Thomson's (2019) English Accent Coach, for example, provides learners with perception practice on English vowels and consonants using HVPT, which he has shown sometimes results in improved production as well. Youglish (2019) is an excellent web-based resource in which students type in any word to hear many different voices saying that word in context, in natural speech. For learners of languages other than English, Forvo.com offers pronunciation of individual words in multiple languages. Hopefully, more resources (including some with a focus on suprasegmentals) in a wide range of languages will be developed.

A desirable tool for designing appropriate software and technology is a naturalistic corpus. Such a corpus is JASMIN-CGN (Cucchiarini, Driesen, Van Hamme, & Sanders, 2008), a repository of contemporary Dutch as spoken by adults, children, seniors, and L2 learners. Orthographic transcriptions, broad phonetic transcription, and part-of-speech tagging have been done. The speech material consists of equal proportions of read and extemporaneous speech. The JASMIN corpus has been used to develop an automatic speech recognition CALL system used for teaching Dutch to immigrants, with a focus on pronunciation. Clearly, the technology and the expertise exist, as the Dutch have shown us, but no similar corpus has been developed for English and many other languages. Corpus development is a huge endeavour and requires teams of people including engineers and applied linguists, but there are promising signs that individuals in the PSLLT community will be working on a multilingual corpus in the near future (Huensch & Staples, 2018).

In 2009, little attention had been paid to the notion of training native speakers (NSs) to be better listeners to L2 accented speech (Goal #6). The only study of which I was aware at that point was conducted by my colleagues and me (Derwing, Rossiter & Munro 2002) in which we trained social work students to better understand accented speech. Although there was a trend in the right direction, no significant improvement in listening pre- and post-training emerged from this study; however, there was a marked improvement in confidence and willingness to communicate (WTC) with L2 speakers, which we saw as an important finding. My colleagues and I have argued that training would be worthwhile for any individuals whose occupations require interaction with members of the public.

Since then, more attention has been given to NS listeners, including some intervention studies. Kang, Rubin and Lindemann (2014) carried out a structured contact activity, based on Kang's dissertation, in which ITAs and undergraduates were brought together to work on a task. Getting to know ITAs in this context led to more positive attitudes towards them as instructors. In another study, Lindemann, Campbell, Litzenberg, and Subtirelu (2016) conducted a short online training program to familiarize NSs with a Korean accent and found modest improvements. Subtirelu and Lindemann (2016) have proposed that native speakers can be helped by (1) improving their attitudes toward nonnative speech, (2) increasing their familiarity with a range of accents and (3) encouraging them to develop and utilize interactional strategies to cope with communication difficulties.

Goal #7 was to eliminate the scapegoating of L2 accents. Problems with pragmatics (knowing what is appropriate to say in a particular context) and grammar can contribute to a lay listener's sense that an accent is to blame. If a speaker uses unexpected phrases or lexical items, they may not be understood, partly because of the general predictability of much of everyday language. Listeners try to anticipate what will be said in a given context, and when something unexpected is produced, they may struggle to understand. Consider this study my colleagues and I carried out (Derwing, Waugh & Munro, forthcoming): we offered 5 weeks of pragmatics instruction on a range of speech acts, including refusals and requests. Intermediate ESL students role-played scenarios with the research team prior to instruction, and then again 5 weeks later. The pre-post scenarios were randomized and played to listeners, who judged the post-test scenarios to be significantly more comprehensible in three out of four cases. The students received practically no pronunciation instruction over this time, but their productions were much easier to understand because of their predictability. L2 pragmatics as an area of study has followed a similar trajectory to L2 pronunciation, and a recognition of the intersection of these two areas is developing (Yates, 2017). Despite the fact that in many language instruction contexts the focus on vocabulary, grammar, reading and writing dominates, I am hopeful that pragmatics will continue to receive more attention.

It has long been known that the more grammar errors an L2 speaker makes, the harsher listeners' judgments will be of their pronunciation (Varonis & Gass, 1982). The implications of this finding for L2 pronunciation have been mostly ignored until recently. Now we have evidence from Lee and Lyster (2018) in a study of French gender that a focus on both grammatical and pronunciation forms will result in better knowledge of grammar and improved comprehensibility. Pronunciation instruction can lead to better use of grammatical forms, and grammatical instruction pointing out the differences between masculine and feminine can lead to better pronunciation. It is clear that researchers are no longer scapegoating accents, but it still happens in the real world.

The “real world” is in need of better strategies for integrating newcomers into their local communities (Goal #7). I will give some examples from Canada, because that is the context with which I am most familiar, but similar strategies may be happening in pockets all over North America and in other immigrant-receiving countries. By no means is Canada doing anywhere near what could and should be done, but the strategies offered here could be implemented almost anywhere.

Dudley (2007) conducted a study in which she surveyed 55 adult ESL students to determine how many of them had done volunteer work to enhance their English. Only 8 had, and only 2 had good experiences. She formulated several recommendations, and a few years later the college where she carried out her research decided to implement them; they developed contacts with ethical businesses and designed a class including a volunteer placement that benefits both employer and student. The employer has additional help (and exposure to L2 learners) and the students receive Canadian experience and references as well as more varied spoken English input than in a standard language class. Both the companies and the students are carefully vetted. This is a very popular program. It requires considerably more work on the part of the teachers than a regular class, of course, but the students find it to be extremely useful on many levels.

Another integration strategy is the Community Connections program, offered by settlement agencies and funded by the federal government. The program matches Canadian-born volunteers with newcomers to facilitate their settlement. Obviously, churches and other community groups can (and do) implement similar programs, but they have to recognize a need in the first place, which may require being approached by an ESL program. We know that the Window of Maximal Opportunity (WMO) for phonological change is in the first six months of massive exposure to English (Derwing & Munro, 2015), so anything that can be done to increase contact with speakers of the local variety of English during that period will be helpful, not only for integration but for pronunciation. Another set of useful resources has been developed by the Toronto Region Immigrant Employment Council (TRIEC); these are geared to issues of inclusion for employers and newcomer employees. TRIEC has freely available videos and other materials that may be useful in other language programs for pragmatics development, which, as noted above, can lead to enhanced comprehensibility.

Finally, another factor to assist the integration of newcomers into their local communities is positive stories in the press and social media. Ray (2018) reports that Peace by Chocolate is a company started by Syrian refugees who came to Canada two years ago and settled in Antigonish, Nova Scotia. The Hadhad family went from making chocolates in their kitchen and selling them at a farmers market to a company that employed 25 people. In August 2018 the family announced that they are hiring 25 more employees to keep up with demand: that is 50 jobs that Antigonish did not have before the Hadhad family came. I chose this story because I happened to read it on a day when I was writing the presentation on which this article is based, but there is a lot of positive press in Canadian mainstream media about the contributions of newcomers, and that helps the overall climate. A majority of Canadians think that immigration is a good thing, although they are worried about irregular asylum seekers (Momani & Stirk, 2017).

Personally, I do not enjoy talking to reporters; indeed, I have colleagues who refuse to do it, because it can be frustrating, but overall, I see it as a chance to get a positive message out. If the media are not covering positive stories in your area, you can approach them yourself with an interesting item about some students, or about pronunciation research that has implications for your community. Typically, news outlets are most likely to be responsive during the summer months and in the December holiday period when they often welcome "soft" news. The more positive the attitudes of the general community members, the more likely they will be willing to talk to newcomers: a plus for integration, language learning, and pronunciation development.

So how close have we come to reaching those Utopian goals set nine years ago? We aren't at Utopia yet. But we have made tremendous progress – so let's keep on trucking!

ACKNOWLEDGEMENTS

I will always be grateful that Murray Munro and I met; we have been fortunate to have a successful, decades-long working relationship that is surpassed only by our friendship. Thanks to John Levis for his foundational contributions to the field of pronunciation research: the PSLLT conference and the *Journal of Second Language Pronunciation*. Lori Diepenbroek, Jennifer Foote, Marian Rossiter, Ron Thomson and Erin Waugh are all former students and current colleagues whose input I value. Bruce Derwing and Ron Smyth have been instrumental in my thinking about L2 pronunciation, and of course, the participation of students, staff, and program directors has been crucial to most of my research. Thanks to the Social Sciences and Humanities Research Council of Canada for generous funding.

ABOUT THE AUTHOR

Tracey Derwing is a professor emeritus of TESL in the Department of Educational Psychology at the University of Alberta, and an Adjunct Professor in Linguistics at Simon Fraser University. With Murray Munro, she has extensively researched L2 fluency and pronunciation, especially the relationships among intelligibility, comprehensibility, and accent. She has also investigated native speakers' speech modifications for L2 speakers, and interventions to enhance native speakers' comprehension of L2 accented speech. Recently, she has conducted workplace studies involving pragmatics and pronunciation. As a director of The Prairie Metropolis Centre for eleven years, Tracey has a strong interest in factors that contribute to successful social integration of immigrants and refugees.

6-102 Education North
Educational Psychology
University of Alberta
Edmonton, AB
T6G 2G5

Canada
780-431-0604
tderwing@ualberta.ca

REFERENCES

Cucchiarini, C., Driesen, J., Van Hamme, H., & Sanders, E. (2008). Recording speech of children, non-natives and elderly people for HLT applications: The JASMIN-CGN corpus. *Proceedings of the 6th International Conference on Language Resources and Evaluation*, LREC 2008, 1445-1450.

Derwing, T. M. (2010). Utopian goals for pronunciation teaching. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 1st Pronunciation in Second Language Learning and Teaching Conference*, Iowa State University, Sept. 2009. (pp. 24-37), Ames, IA: Iowa State University.

Derwing, T. M. (2018). The efficacy of pronunciation instruction. In O. Kang, R. Thomson & J. Murphy, (Eds.). *The Routledge handbook of English pronunciation*, (pp. 339-353). New York: Routledge.

Derwing, T. M. (2018). The role of phonological awareness in language learning. In P. Garrett & J. M. Cots (Eds.). *The Routledge handbook of language awareness* (pp. 339-353). New York: Routledge.

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

Derwing, T. M., Rossiter, M. J., & Munro, M. J. (2002). Teaching native speakers to listen to foreign-accented speech. *Journal of Multilingualism and Multicultural Development, 23,* 245-259.

Derwing, T. M., & Waugh, E. (2012). Language skills and the social integration of Canada's adult immigrants. *IRPP Study 31*. Montreal: Institute for Research on Public Policy.

Derwing, T. M., Waugh, E., & Munro, M. J. (forthcoming). Pragmatically speaking: Preparing adult ESL students for the workplace.

Dudley, L. (2007). Integrating volunteering into the adult immigrant second language experience. *Canadian Modern Language Review, 63,* 539-561.

Grant, L. (2014). *Pronunciation myths: Applying second language research to classroom teaching.* Ann Arbor: University of Michigan Press.

Grantham O'Brien, M. G., Derwing, T. M., Cucchiarini, C., Hardison, D. M., Mixdorff, H., Thomson, R. I., Strik, H., Levis, J. M., Munro, M. J., Foote, J. A., & Muller-Levis, G.

(2018). Directions for the future of technology in pronunciation research and teaching, *Journal of Second Language Pronunciation, 4,* 182-206.

Huensch, A., & Staples, S. (2018). Towards a protocol for a multilingual corpus for pronunciation researchers. Invited colloquium, PSLLT 10, Sept. 8, Ames, Iowa.

Isaacs, T., & Trofimovich, P. (2012). Deconstructing comprehensibility: Identifying the linguistic influences on listeners' L2 comprehensibility ratings. *Studies in Second Language Acquisition, 34,* 475-505.

Kang, O., Rubin, D., & Lindemann, S. (2015). Mitigating US undergraduates' attitudes toward international teaching assistants. *TESOL Quarterly, 49*, 681-706.

Kermad, A., & Kang, O. (2018). Effect of classroom assessment stakes on English language learners' oral performance. *TESOL Journal*, Advance view, https://doi.org/10.1002/tesj.392

Lee, J., Jang, J., & Plonsky, L. (2015). The effectiveness of second language pronunciation instruction: A meta-analysis. *Applied Linguistics, 36*, 591-610.

Lee, A., & Lyster, R. (2018). The missing link in the acquisition of grammatical gender by second language learners of French. Presentation at AAAL, Chicago, March 24.

Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly, 39,* 369-377.

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

Levis, J., & Grant, L. (2003). Integrating pronunciation in to ESL/EFL classrooms. *TESOL Journal, 12,* 13-19.

Lindemann, S., Campbell, M. A., Litzenberg, J., & Subtirelu, N. C. (2016). Explicit and implicit training methods for improving native English speakers' comprehension of nonnative speech. *Journal of Second Language Pronunciation, 2*(1), 93-108.

Loewen, S., & Isbell, D. R. (2017). Pronunciation in face-to-face and audio-only synchronous computer-mediated learner interactions. *Studies in Second Language Acquisition, 39,* Special Issue 2, 225-256.

Momani, B., & Stirk, J. (2017). *Diversity dividend: Canada's global advantage.* Waterloo, ON: Centre for International Governance Innovation and Fondation Pierre Elliott Trudeau Foundation.

Murphy, J. (2013). *Teaching pronunciation.* Alexandria, VA: TESOL.

Pickering, L. (2018). *Discourse intonation.* Ann Arbor: University of Michigan Press.

Ray, C. (2018). Syrian chocolate company in Antogonish set to double in size. CBC News, August 5. https://www.cbc.ca/news/canada/nova-scotia/peace-by-chocolate-antigonish-company-doubling-in-size-1.4774576

Ruivivar, J., & Collins, L. (2018). The effects of foreign accent on perceptions of nonstandard Grammar: A pilot study. *TESOL Quarterly, 52*, 187-198.

Subtirelu, N., & Lindemann, S. (2016). Teaching first language speakers to communicate across linguistic difference: Addressing attitudes, comprehension, and strategies. *Applied Linguistics, 37,* 765-783.

Thomson, R. I. (2019). English Accent Coach. www.englishaccentcoach.com.

Thomson, R. I. (2018). High variability [pronunciation] training (HVPT): A proven technique about which every language teacher and learner ought to know. *Journal of Second Language Pronunciation, 4,* 207-230.

Thomson, R. I., & Derwing, T. M. (2015). The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics, 36,* 326-344. doi:10.1093/applin/amu076.

Toronto Region Immigrant Employment Council (TRIEC). (2018).TRIEC website. Accessed April 27, 2012. http://triec.ca/employers/online-learning/

Youglish. (2019). https://youglish.com.

Varonis, E., & Gass, S. (1982). The comprehensibility of nonnative speech. *Studies in Second Language Acquisition, 4,* 114-136.

Yates, L. (2017). Learning how to speak: Pronunciation, pragmatics and practicalities in the classroom and beyond. *Language Teaching*, *50*, 227-246.

Dickerson, W. (2019). The ripples of rhythm: Implications for ESL instruction. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching conference*, ISSN 2380-9566, Iowa State University, September 2018 (pp. 36-54). Ames, IA: Iowa State University.

INVITED TALK

THE RIPPLES OF RHYTHM: IMPLICATIONS FOR ESL INSTRUCTION

Wayne B. Dickerson, University of Illinois at Urbana-Champaign

The work of Brazil, Coulthard, & Johns (1980), Bolinger (1986), Cauldwell (2002), Wells (2006), and others has led to a growing consensus about spontaneous English phrases: Their rhythm consists predominantly of only one pitch accent (nucleus) or two pitch accents (onset and nucleus) that alternate with unaccented syllables. If we accept these findings, then our pronunciation teaching will differ from our traditional TESOL approach that has been so profoundly shaped by Prator's (1951) version of stress timing. This paper explores the implications of this radically different model of rhythm for the content and presentation of ESL/EFL pronunciation instruction.

INTRODUCTION

Most ESL instructors would agree that we teach pronunciation to help learners communicate intelligibly in oral English. To achieve this goal, many types of language content go into the curriculum. If we consider only the pronunciation part of such a curriculum, as represented by recent pronunciation textbooks, the range of relevant topics is amazingly wide. The prospect of covering so many areas in whatever time is available is truly daunting. In decades of teaching pronunciation, I have lost track of how many times I have dropped, added, and reordered topics in an effort to create a course with the most effective and well-motivated mix of content for my students.

This creative process has led me to the conclusion that there is only one pronunciation topic that rises to the top of a list of priorities because of its centrality to the goal of helping learners communicate intelligibly in oral English. The topic is rhythm. I give it top billing because I believe it alone has the potential to organize the content of a pronunciation course in a way that keeps instructors' and learners' attention on the stated goal of intelligible oral English.

To support this claim about the potential of rhythm, I start by identifying what I mean by rhythm and why it is pivotal to any effort to improve the intelligibility of learners' oral communication. We will then be in a position to consider the implications of adopting and promoting this model: What have we been doing well in our teaching of rhythm, and where and why have we missed opportunities to offer important guidance to our students?

THE TWO-PEAK PROFILE

Lucy Pickering (2018) summarizes David Brazil's model of the rhythm found in spontaneous English:

> It can be useful to think of the typical structure of a tone unit in English as comprising three to seven words and containing one or two prominences. (p. 23)

Pickering's summary applies equally to the conclusions of other researchers like Kenneth Pike (Fries 1945, p. 64), Dwight Bolinger (1961, p. 135), Richard Cauldwell (2013, p. 39), and John Wells (2006, p. 192), whose work attempts to characterize the language found in naturally occurring talk.

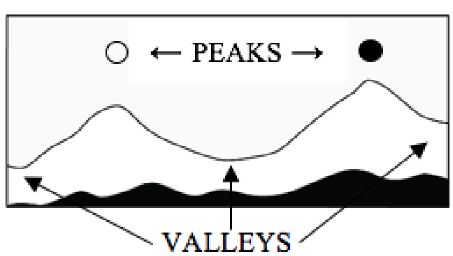
This summary does not describe stress-timed rhythm, a linguistic hypothesis about how English rhythm works that, when repeatedly tested by phoneticians, was found to be universally unsupported (Dauer, 1983; Roach, 1982). (See Cauldwell (2002) for an overview of the topic and Arvaniti (2012, p. 351ff) for a detailed examination of the many studies that tested the hypothesis.)

Clifford Prator, Jr. (1951) introduced the model of stress timing to ESL/EFL instructors through his ESL pronunciation textbook. However, he did so with a simplification of his own that was foreign to the original linguistic hypothesis. He taught his students to use a pitch accent on **every** content word in a phrase. It is this now-discredited model in a distorted form that has become the de facto standard in our language-teaching field and now appears in almost every ESL/EFL pronunciation text, teacher's guide, and teacher-preparation course (Dickerson, 2015).

Given the centrality of rhythm to intelligible oral communication, as argued in this paper, there could not be a worse place in phonology to distort the linguistic facts. Since rhythm is a feature of every spoken phrase, stress-timed rhythm yields a form of oral English that no one speaks natively. For more than half a century, with no malice of purpose—indeed with every good intention, our profession has nevertheless been teaching students around the world a non-English (and often unrecognizable) way to communicate. ESL/EFL learners who have acquired an ability to speak English well have done so despite the instruction they have had in rhythm.

By contrast, the model of rhythm advanced here is one that has evolved partly from the fieldwork of phoneticians examining spontaneous speech. It reflects the actual rhythm of English. Just as rhythm is the worst place in phonology to distort the linguistic facts about so basic a feature, it is also the best place in phonology to get the facts right for the sake of our students and their communicative effectiveness.

To elaborate on this model of rhythm in spontaneous speech, I want to introduce the pedagogical terminology I use when teaching ESL/EFL students (and will use in this paper). To start with, since the model had no name—no counterpart to "stress-timed rhythm," I drew on the metaphor of a mountain range in profile and called it the **two-peak profile.** The name emphasizes peak-valley alternation, not timing, as the fundamental nature of English rhythm.



In our pronunciation materials, we refer to the second pitch accent as the **primary peak,** or simply the **primary**. It is commonly referred to as the **tonic** or the **nucleus** in linguistic literature. We mark this peak with a filled bullet over the accented vowel. The first pitch accent, if there is one, is the **anchor peak**, or simply the **anchor**. In phonetics, it is known as the **onset** or **onset syllable** (Wells, 2006, p. 8; Tench, 1996, p. 135). We mark this peak with a hollow bullet over the accented vowel. Following are examples of the two-peak profile from speech samples.

(1) Have you had lunch yet?

(2) Twenty twenty's a leap year.

(3) Usually he won't comment on them.

(4) I bought these peaches in Savannah yesterday.

Before, between, and after these peaks are what we call **valleys**—syllables that carry no pitch prominence. Sometimes, to draw special attention to valleys, we mark them with subscripted swooshes, as shown above. But even without swooshes, all non-peak syllables are deaccented. There are potentially five positions in the two-peak profile: two peaks and three valleys. While there must be a primary peak, all other parts are optional. For instance, example (1) has no anchor nor medial valley. Example (3) is missing an initial valley. Examples (2) and (4) have all five parts.

Although the two-peak profile has a long history in phonetics (see Pike in Fries, 1945, p. 64; Bolinger (1961, p. 135), we think of it as new because it is largely unknown in TESOL circles. But not entirely so. It first appeared in David Brazil's pedagogical materials published in 1994. More recently, Gorsuch, Meyers, Pickering, and Griffee built on Brazil's work in their *English Communication for International Teaching Assistants* (2013). These are pioneering efforts, but resistance to an alternative model of rhythm is formidable; the old model of stress-timed rhythm remains strongly entrenched in our field. Furthermore, over the past three decades, it has been hard to get excited about a model for which no full rule had emerged to predict the first peak.

Is there a rule for where the anchor occurs? We have rules for nearly all of phonology—for placing the primary peak, for how to stress multiword numbers, for how final consonant clusters are simplified, for why an intonation pattern rises or falls at the end of a phrase, and so on (Dickerson, 2004). Given that language is profoundly regular, a rule must also exist for the anchor. A stronger case for assuming that an anchor-placement rule exists is that native speakers of English behave in a rule-governed way: They can tell when an anchor is used neutrally—without drawing special attention—and when it is used emphatically (Dickerson, 2015).

The need for this rule is felt acutely by adult learners who want to take responsibility for their learning and do not want to depend on teachers as crutches for where to put pitch accents (Dickerson, 2004). They need an anchor-placement rule to use in their private self-practice to check and correct their own accuracy just as they need other pronunciation rules. Furthermore, given how commonly we use emphasis in everyday language, learners also need the rule by which to determine when a co-speaker is speaking non-emphatically and emphatically. These learners also need the rule in order to know where to place anchors with the meaning they intend.

To address this felt need and to encourage greater interest in and adoption of the two-peak profile, in 2013 my co-author, Laura Hahn, and I set our sights on working out the rule for the dominant uses of the anchor, translating it into a pedagogical form, and trialing it with classes of university-level students and international teaching assistants. The second edition of our textbook, *Speechcraft: Discourse Pronunciation for Academic Communication* (in press, University of Michigan Press) is built around the two-peak profile and accent-placement rules for the primary and anchor peaks.

THE LINCHPIN LINKING SOUND TO MEANING

Intelligible spontaneous speech—the goal of pronunciation teaching—is speech that the **listener** can understand in the rapid give-and-take of conversation. Since the listener is the judge of what is intelligible, what must a spoken phrase contain to make it so? The answer that research gives is that the phonetic cues to the core message—the **focus**—must stand out to the listener (Wells, 2006, p. 234). The listener, in turn, must know how to decipher the phonetic cues to discover the focus if the communication is to be successful.

Focus is a term that has found its way into many pronunciation textbooks (Gilbert, 2012; Gorsuch, et al. 2013; Grant, 2017). Originally synonymous with the word carrying the primary peak, it meant what is new or in contrast (Ladd, 1996, pp. 225ff). Although we now understand that English rhythm has one or two peaks per phrase, the concept of focus is still tethered to peaks. In the case of two peaks, it now starts with the cue in the anchor word and ends with the cue in the primary word. In the case of one peak, the cue for the start and end of the focus is in the same primary peak. To refer to both cases, Wells uses the term *focus domain* (2006, p. 116), and Brazil et al. prefer the term *tonic segment* (1980, pp. 39-42).

To foreground the focus of a phrase for the listener, the speaker uses pitch accents—the peaks—of rhythm to identify the words and constructions that are most germane to the message. The focus itself is not in the phrase but is instead in the pragmatic interpretation of the phonetic cues in the phrase.

(1) Have you had lunch yet? (2) Twenty twenty's a leap year.

For example, the speaker's peak in (1) should tell the listener that the heart of the question is *lunch*, not the suprasegmentally enhanced vowel in *lunch*. Similarly, in (2), the speaker's pitch accents on the main vowels in the second *twenty* and in *leap* should lead the listener to interpret the focus as the multiword number *twenty twenty* and the compound noun *leap year*.

I find the current definition of the focus, namely, the cue(s) in the anchor-word-through-the-primary-word string, particularly valuable conceptually and pedagogically because it identifies the part of a phrase that the listener must understand in order for the phrase to be intelligible. This string allows us to describe the essence of the message better than using the primary peak alone. In fact, for pedagogical purposes, we prefer to call this string the **essence** or **semantic essence** of the phrase rather than the focus.

Since noticeable pitch accents are essential to a listener's grasp of the speaker's message, they should also be essential to what we teach learners to produce and interpret. That is, the objective of any course that aspires to help learners communicate intelligibly in spoken English should be the clear production of the two-peak profile and the facile discovery of the focus in that profile.

The reason we assign this rhythmic profile the highest priority in pronunciation instruction is its facilitative role in communication and in teaching the sound system. First, its peaks highlight the part of a phrase—the focus—that most directly signals the speaker's message and that is most relevant to the listener's understanding of that message. Second, and equally important, rhythm is the gravitational center of the English sound system; the entire phonology is concentrated on implementing that rhythm. In fact, much of phonology is dependent on the precedence of rhythm. For example, as critically important as intonation is to the intelligibility of the focus, its pitches hang on the peaks of rhythm.

IMPLICATIONS FOR INSTRUCTION: PERSPECTIVE

What are the consequences that flow from adopting the two-peak profile for our ESL/EFL instruction? Some consequences affect how we present pronunciation instruction to learners.

1. A well-motivated starting point. If the reason we teach pronunciation is to help learners make their phrases easy for listeners to understand by meeting their expectations, then instructors should state this objective at the start of each semester. Furthermore, if the way to reach this objective is to teach learners to pronounce what listeners are listening for, namely, the focus signaled in the two-peak profile, then it stands to reason that instruction on the two-peak profile must start each semester, as well.

Rhythm as a starting point is novel to most learners because it has not been part of their prior ESL/EFL learning. Rhythm also has a nebulous quality because it does not seem so concrete as other parts of the sound system, like consonant and vowel segments or lexical stress. Even intonation seems easier to grasp than rhythm. Rhythm is all the more startling because the claims we make for it are so consequential, namely, its centrality to the intelligibility of every spoken utterance. How could something so relatively unknown in learners' experience with English and so seemingly intangible really be so important? There is shock value in this starting point that, for many, is actually intriguing. It says: "This instruction will be different from what you may have expected. Nevertheless it will turn out to be really important to your success with English."

2. A consistent, unifying rationale for instruction. An overarching benefit of recognizing that all phonology works toward the single objective of creating noticeable pitch accents is that, as we take up each new topic, we can relate the relevance of the new topic to the master objective of creating a rhythm that communicates the focus of the message. By returning again and again to rhythm as the point of reference for every new topic, we create a thematic unity of subject matter that would otherwise seem like disparate, unrelated parts of the sound system.

While some students accept each topic with equanimity as it comes along, we have found that most learners feel decidedly better about studying a topic when they know how it affects their personal communicative effectiveness. Instructors, too, have a stronger sense of mission when they articulate the connection between what they teach in each lesson and the goal of their instruction, namely, intelligible communication achieved through an effective rhythm.

IMPLICATIONS FOR INSTRUCTION: WHAT, WHEN, AND HOW WE TEACH

Other implications for teaching the two-peak profile go beyond orientation and motivation. They entail adding and reorganizing content and teaching with new techniques and emphases. What have we done well in the area of rhythm, and where are the gaps we could fill to prepare our students even better to use the two-peak profile in their interactions with English speakers?

3. The role of grammar in making pronunciation decisions. The terms **content words** and **function words** (Pike, 1945, p. 118) are staples in discussions of pronunciation. Since these categories are defined by parts of speech, they tell us that grammar is fundamental to pronunciation choices. This should come as no surprise. Since language is for communicating meaning, both grammar and pronunciation point toward the same goal. Grammar does it by arranging parts of speech; pronunciation does it by arranging sounds, stresses, and pitch.

The traditional, although tacit, assumption about the convergence of grammar and pronunciation is that the two map onto each other so well that basic parts of speech are adequate to describe and predict how the sound system works. ESL/EFL learners' background in grammar should therefore be a satisfactory preparation for pronunciation work; it should be enough that students can recognize, for example, content words as nouns, adjectives, verbs, and adverbs.

OLD (and inadequate) goal: Use general categories of grammar to predict pronunciation.

The reality is that the match between parts of speech and pronunciation choices, while close, is not one-to-one. There is no part of speech that pitch accents prefer categorically. As examples, look at a sampling of anchor placements among content words. In (a) the anchor is on the first noun, but in (b) it is not; the first noun is in a valley. In (c) the anchor is on the first adjective, but in (d) it is not; the first adjective is in a valley. In (e) the anchor is on the first verb, but in (f) it is not; the first verb is in a valley. In (g) the anchor is on the first adverb, but in (h) it is not; the first adverb is in a valley. An ability to recognize general parts of speech is, unfortunately, not sufficient to understand nor to predict the behavior of the anchor peak.

(a) My roomie's carrying too many courses.
(b) My roomie's course load is too heavy.
(c) Your best option's to take it online.
(d) Green tomatoes are great when grilled.
(e) Explain the implications of your data.
(f) She believes her data's quite compelling.
(g) He's certainly dull as a lecturer.
(h) He's rather dull as a lecturer.

Often the grammatical function of a part of speech is germane to a pronunciation choice. For example, the subject noun carries the anchor in (a), but the possessive noun in (b) does not. This point leads us to a more adequate goal for preparing learners for pronunciation work.

NEW goal: Teach parts of speech with enough detail to make good pitch-accent decisions.

Accordingly, in our instruction, we ask students early in the semester to review parts of speech and their functions. It is not necessary to brush up on all of English grammar when the review is specifically tailored to pronunciation needs. This review (an appendix in *Speechcraft*) is done out of class as self-study and followed up with an assessment of learners' grammatical readiness for the study of pronunciation.

Not only do students perform better after doing a part-of-speech review, but the targeted study also helps to level the playing field for class members. Those whose formal exposure to English grammar is only marginally adequate are not disadvantaged by their weak background, and even those who have a good grasp of English grammar refine their understanding.

4. An identifiable language style with its own characteristics. Few pronunciation texts are explicit in identifying the target style of English they present. Even fewer attempt to teach the features of a particular style. Perhaps the assumption is that details of style do not matter as long as the language taught is educated speech.

OLD (and inadequate) practice: Teach English pronunciation without specifying a style and its features.

The problem comes down to one’s definition of educated speech. Judging from current pronunciation texts, educated oral English is not dominated by the two-peak profile. It puts no constraints on phrase length. Nor does it require compression devices in production. This is not what listeners wish to hear when spoken to. Their choice of style is the one they themselves make when speaking, namely, spontaneous educated speech. This style is characterized by (a) the familiar two-peak profile, (b) short phrases, and (c) an abundance of valley compression.

As we are coming to understand, this trio of spontaneous-speech features promotes intelligibility. Early sections of this paper described the centrality of the two-peak profile to identify for listeners cues to the beginning and end of the focus. The rhythmic profile, however, does not perform its pragmatic function alone; it requires the help of two other features of spontaneous speech—constraints on phrase length and pervasive compression phenomena—both of which address the same issue, namely, the limitations of the listener’s memory.

The listener’s goal is to snatch the essence of a phrase out of the air as the speaker talks. The target of the listener’s attention is everything from the anchor word through the primary word, including any in-between valley words. It is the in-between stretch that can tax memory because it makes a difference to the listener how much time the speaker takes to get from the anchor word to the primary word. For the sake of memory, the shorter the better, lest the listener forget the first peak before the second arrives, slowing comprehension (Kjellin, 1999; Munro & Derwing, 1998). Anything that shortens inter-peak time benefits memory.

A **short phrase length** helps memory because the closer the anchor and primary peaks are to each other, the quicker the focus can be revealed to the listener. **Valley compression** also helps memory because it directly minimizes time spent in the valley between the anchor and primary peaks. Each of these features deserves attention. Let us consider first how we help learners create well-sized phrases. The topic of compression comes in the following section.

Although English grammar does not limit the length of a spoken phrase, cognitive factors like the capacity of speakers to take in, and listeners to process, speech on the fly do impose limits. The phrase-length solution to the problem of retaining a memory of the whole focus is to keep the anchor and primary peaks close together by admitting at most only five words between them. When the maximum number of words in a phrase is held to seven, inter-peak time is controlled.

While limits on phrase length come naturally when speaking spontaneously, limits are not automatic for learners who are practicing with a written text. Unaware of how short spoken language phrases really are, learners need guidance. We offer it to them by starting with an example of spontaneous speech to make three points. First, phrases are typically no longer than seven words and average only three to four words. Second, pauses between phrases occur at identifiable places that can be discovered by examining spontaneous speech. Third, to speak intelligibly, learners must practice using phrase-boundary guidelines to convert written language into phrases that simulate spontaneous speech. These three points are not customarily part of ESL/EFL pronunciation instruction despite their being necessary to intelligible speech.

To model appropriately sized phrases, all instructional materials should conform to constraints on phrase length so that students continue to learn by example. Then, whenever they compose materials to be spoken, they will be more likely to remember to adhere to good models and to practice written-to-spoken conversion techniques. The point we emphasize is that when spoken phrases are sized correctly, the two-peak profile fits more comfortably on them, and listeners get the speaker's point more readily, than when phrases are overly long and strain listeners' memory and processing capabilities.

NEW goal: Adopt spontaneous speech as the oral target; teach written-to-spoken conversion techniques to simulate spontaneous phrases; practice the two-peak profile with short phrases.

Of course not all talk is spontaneous, particularly in academic and professional settings where presenting papers and proposals is expected. I strongly believe that the best preparation for formal speaking is learning to converse well in everyday interactions. The characteristics of extemporaneous speaking are basic to good public speaking—good peak-valley contrasts in the two-peak profile on short phrases that exhibit comfortable compression in their valleys.

5. Natural Speech Phenomena. Spontaneous speech is characterized not only by the two-peak profile and by short phrases but also by the pervasive use of compression devices (assimilation, trimming, reduction, and linking). These devices provide another solution to the listener's challenge of catching and remembering the whole focus of a phrase in the rapid flow of speech. By miniaturizing each valley syllable to the extent possible, the speaker can quickly deliver to the listener cues to the focus in a single continuous string. In turn, the listener is more likely to recognize the focus and take it in fully when it comes in the context of a familiar rhythm used in short phrases (Kjellin, 1999, pp. 23f).

In the repertoire of pronunciation topics, natural speech phenomena were among the last to be added to pronunciation textbooks. Often relegated to a lesson late in the text, the emphasis was on the compression of function words (Dale & Poms, 1994). The mistaken impression, arising from stress timing, is that all content words will carry a pitch accent and therefore require no compression.

From the perspective of the two-peak profile, we now understand that every word that can contribute a peak to the two-peak profile can also contribute its syllables to valleys. Even so, compressed speech has not yet achieved parity with other pronunciation topics. Two issues stand in the way.

First, a widely accepted conclusion about natural speech phenomena is that we must teach these devices mainly for purposes of perception (Levis, 2018, pp. 148-149). Learners' struggle to understand an incoming speech stream that has been reshaped by compression devices is painfully obvious, and research has confirmed their problem. But to downplay learners' responsibility to produce such a speech stream for listeners is to say that the intelligibility of the learners' speech is not important to listeners. This is the opposite of the goal of pronunciation instruction, as stated in the first point above, namely, to help learners make their phrases easy for listeners to understand by meeting listeners' expectations.

OLD (but incomplete) goal: Teach learners to understand compressed speech but not to produce it.

When we elevate learners' difficulty to understand a speaker's compressed speech but dismiss listeners' frustration trying to understand learners' uncompressed speech, we expose our bias and belie our belief that intelligibility is a reciprocal requirement for speakers and listeners alike.

Second, although natural speech devices are part of all modern pronunciation texts (Miller, 2006), they are typically taught as discrete topics to improve the naturalness and flow of speech, not as an integral part of a coordinated effort to make it possible for the listener to grasp the focus in one take. A better way to approach these devices is to understand them as contributing to the listener's success at understanding the core meaning of a phrase. This means introducing these devices together as soon as learners begin producing the two-peak profile, namely, early in instruction. From a practical point of view, how does it help learners to use the rhythm model, introduced early in instruction, if consonant and vowel compression comes late in instruction or is scattered among learners' pronunciation lessons?

NEW goal: Teach consonant and vowel compression techniques in connection with the two-peak profile as equally important for perception and production.

6. Rhythm as a pacing tool. A myth about English that we hear repeatedly from our students is that fluent speech is fast speech. This assumption leads some to try to say everything as quickly as they can. Despite the frequency of this error, it is rare that pronunciation texts comment on pacing one's oral delivery except to urge students to hurry their articulation of function words.

OLD (and inadequate) practice: Speed up function words; slow down content words.

In point of fact, every language is delivered at a range of speeds (Cauldwell, 2013, pp. 94ff). The belief that a target language is spoken especially fast reflects more the learners' inability to process it as quickly as it is spoken than it does the actual articulation speed of the new language.

A more helpful rebuttal to the myth of speed is this fact: No single speed is appropriate for all situations, nor is it appropriate for all parts of an English phrase. Peak syllables sound slow when they are stretched out, and valleys syllables sound fast when they are compressed. Hurrying and slowing are rhythm skills. They need to be practiced in conjunction with the two-peak profile because it serves as a traffic signal for when to speed up and when to slow down in a phrase. When the signal is obeyed, each phrase is delivered with speeds expected of intelligible speech.

NEW goal: Include in a description of the two-peak profile its use to control the speed of speech.

7. Switching patterns: The Heart of the Two-Peak Profile. A two-peak profile has, at most, two valley-to-peak changes in prominence (first-valley-to-anchor and middle-valley-to-primary) and two peak-to-valley changes in prominence (anchor-to-middle-valley and primary-to-last-valley). We call these **switching patterns** because they are sudden flips in prominence that often happen from one syllable to the next. Without these two rapid switches—peak-to-valley and valley-to-peak—there would be no two-peak profile.

In general, this point sounds like something from stress-timed rhythm instruction: For decades, pronunciation teachers have encouraged their students to put peaks on content words and push function words into valleys. The idea of contrast is not new, nor is it wrong.

OLD (inadequate) goal: Contrast peaks (on all content words) with valleys (on function words).

However in two respects, making a difference between content words and function words is not enough for the two-peak profile. Roger Kingdon (1958, p. 160) identifies one issue: “The difference in prominence between stressed and unstressed syllables is greater in English than in many languages.” That is, many students will find that the degree of peak-valley difference needed for English switching patterns lies outside their customary range. When speaking English, their peaks must be more distinct from their valleys in pitch, duration, and intensity.

The second issue becomes clear with some actual examples. Here are two valley-peak switches:

And here are two peak-valley switches:

Learners accustomed to stress-timed rhythm phrases may be startled to see **whole** content words in valleys, such as the words *bought* and *yesterday* (see the stars). The vowel in each syllable of these content words must be suppressed.

NEW goal: Monitor the size of contrasts and use of valley vowels for content words in valleys.

An attention to the magnitude of switches and to the suppression of content-word vowels is still not enough to prepare learners to use vowels fully to implement a rhythm that communicates clearly to listeners. The story of vowels continues.

8. The Prosody of Vowels. We have noted the traditional emphasis on developing learners’ skill at distinguishing phonemes such as /iy/ and /ɪ/ so that listeners can hear the difference between words like *reach* and *rich, cheap* and *chip.* We have also highlighted the importance of including work on vowel compression to reduce the time of valleys. Our focus in both cases is on guiding learners’ control of the **articulatory** features in the lower half of the vowel circle below.

Kingdon's (1958, p. 160) caution about the need for adequate contrast between peaks and valleys points us to another area of vowel work that is generally missing from pronunciation instruction. When we contrast a peak with a valley and vice versa, what part of a syllable do we change? The answer is the **vowel** part. But what aspect of vowels do we change when we switch from a peak to a valley and vice versa? We do not manipulate their articulatory features but their **prosodic** features—those in the upper half of the vowel circle.

Contrast, as in a switching pattern, is the exclusive domain of vowel prosody. Whether we use the two switching patterns to create word rhythms around a single peak (peak-valley, valley-peak-valley, or valley-peak) or phrase rhythms around potentially two peaks, we do so by adjusting the duration, pitch, and intensity of vowels. Although they create different rhythm patterns, switching patterns at these two levels of structure are identical. This is why it makes sense to teach switching patterns at the word level; they transfer directly to the phrase level. That is, a peak at the word level can create one of the peaks at the phrase level. Of course, there are no switching patterns in words found in phrase-level valleys because there are no peaks there.

Since prosody-based switching patterns using vowels are the means by which a speaker tells a listener where the focus of the phrase begins and ends, I was obliged to adjust my goal for vowels once more, this time to include vowel prosody.

NEW goal: Teach vowel articulation and v*owel prosody* together.

Since manipulating prosody is new to vowel instruction, some example exercise types may be helpful. As noted, the place to begin to meet this expanded goal is at the word level. We work on the **perception** of prosodic vowel differences first by asking learners to listen to a pair of vowels in adjacent syllables. Here is an exercise on the diphthong in *now*. Of course learners are naturally drawn to the most noticeable syllable. To encourage them to pay attention to the full range of vowel prosody, we ask them to pick out the least noticeable syllable. (More items than these would be used with learners, of course.)

> I'll say a word twice. Identify its quietest syllable by saying 1st or 2nd. For example, if I say ***óutbòund***, ***óutbòund,*** you would say ***2nd*, for the second syllable**. But if I say ***bòw dówn***, ***bòw dówn***, you would say ***1st*, for the first syllable**.

óutbòund,	*2nd*	òutshóut,	*1st*
bòw dówn,	*1st*	tównhòuse,	*2nd*
lóudmòuth,	*2nd*	fòuled óut,	*1st*
chòw dówn,	*1st*	cóuntdòwn,	*2nd*

Next we work on the **production** of prosodic vowel differences at the word level. The focus is on pronouncing an adequate difference of the same vowel. If the target vowel is the same as above, then we can use items from the perception exercise for production purposes. We monitor for articulation and for prosodic contrast. For example, is the first vowel in *loudmouth* sufficiently different in its prosodic features from the second vowel to register as a peak-valley contrast?

Each time we introduce a vowel, we include perception and production exercises that offer learners practice identifying and modulating vowel prosody.

From isolated words and constructions (e.g., *cóuntdòwn, fòund óut*) we move to contextualized phrases but still contrasting the same vowels. Here is an excerpt from a dialogue about a student who was in a traffic accident (Dickerson & Hahn, in press). It illustrates how peak-valley and valley-peak switching patterns using vowel suprasegmentals fit into and promote the two-peak profile. Two instructors are talking.

> A: Did you hear about Maria? She was in a traffic accident.
>
> B: I'm sorry to hear that. What happened? How is she?
>
> A: She was walking along Peach Street, | composing a text message, | and strayed into a bike lane. She collided with a cyclist | and came home with a few bruises | but no broken bones. That was three weeks ago.

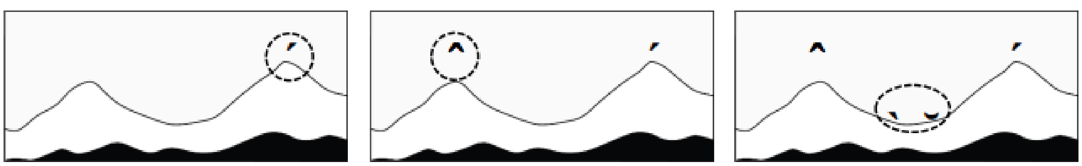
Instead of starting with this dialogue as an interaction, we start with word pairs from the dialogue where the same vowel is used in both words, creating a peak-valley switch and a valley-peak switch between a primary-peak vowel and a lesser-stressed vowel. That is, we again hold the vowel constant so students can focus their attention on the change in prosody.

Peak-Valley Contrasts	**Valley-Peak Contrasts**
● [She was in a] traffic accident /æ - æ/	○ ● [came home with a] few bruises /uw - uw/
○ ● [She was walking along] Peach Street /iy - iy/	○ ● [but no] broken bones /ow - ow/
○ ● [composing a] text message /ɛ - ɛ/	○ ● [That was] three weeks [ago] /iy - iy/

After this kind of practice, students use the dialogue itself. Contrasts at the level of two-word phrases are the kind that can lead to a clear rendition of the two-peak profile in the dialogue.[i]

9. Contrasts with Word Stress. Contrasts in vowel prosody are, of course, also contrasts in word stress. However, by starting with vowels in words, a smaller context, we can isolate prosodic issues with the components of stress—pitch, intensity, and duration—and deal with them more easily than we can with word stress in phrases, a larger context. This is the ideal environment to work with those students whose native languages have left them less sensitive or insensitive to the prosodic features that are relevant to English stress (Levis, 2018, pp. 109-114).

Before continuing with word stress, let us change the notation system from pedagogical bullets to conventional stress marks, which allows us a more refined look at word stress. By using the quaternary stress-marking system (´ ˆ ` ˘), we can represent the stress of the two-peak profile as well as the stress of words. Since there are different stress-marking systems, note how the quaternary stress marks are used instead of bullets.



PRIMARY PEAK ANCHOR PEAK VALLEY STRESSES

For the two-peak profile, the acute stress mark is used for the vowel that carries the primary peak, as *lunch* does in *Have you had lúnch yet?* The circumflex stress mark is for the vowel that carries the anchor peak, like the first vowel of *usually* in *Ûsually he won't cómment on them.* These are the two levels we call **maximized** vowels. Valley syllables are also of two kinds. Those that have a lightly stressed vowel are marked with a grave stress mark, and those that have an unstressed vowel are marked with a breve stress mark. Lightly stressed vowels are those in the words *Ì*, *bòught,* and *thèse,* and unstressed vowels are those in *-chĕs ĭn Săv-* and *-ă* in *Ì bòught thèse pêachĕs ĭn Săvánnă.* These are the two levels of **minimized** vowels.

For the citation form of a word, as found in a dictionary, only three levels of stress are needed. Since a multisyllabic word has only one heavy stress, there is a primary peak but nothing comparable to the anchor peak at the word level; no circumflex stress mark is used at this level. All non-primary peak syllables are valley syllables, containing either lightly stressed vowels indicated with a grave stress mark or unstressed vowels indicated with a breve stress mark.

Resuming our topic of word-stress, what do we teach when we teach word stress? The typical answer to this question is that we teach the **position** of the main stress in a word. For example, is the main stress of *fantastic* on the first syllable, the middle syllable, or the last syllable? If my students are able to hear, predict, and produce the heavy stress of a word on the right syllable, I applaud their success. They have accomplished something important that transfers directly to one aspect of rhythm: The primary peak is positioned on the main stress of a multisyllabic word. So the position of the main stress in a word is a worthy part of what we want our students to learn. It is also content typically found in modern pronunciation texts.

OLD (but incomplete) goal: Teach the position of the main stress in a word.

However, when examined from the perspective of our end goal of helping students communicate intelligibly, this objective is incomplete. It says nothing about how the focus of a phrase is signaled in rhythm. By stopping with the position of the main stress, I missed another connection between word stress and phrase rhythm. In effect, I was no better than a dictionary.

Dictionaries tell learners that the main stress of *popularity* is on the third syllable. In citation form, that stress is marked as a primary stress—*populárity*—because the citation form of a word is a one-word phrase with a primary peak on the main stress. Is the primary stress appropriate everywhere in the two-peak profile? Of course not. It is appropriate when the word carries the primary peak, but it is not appropriate for the anchor because we drop the level of stress for the anchor. A primary stress is certainly not appropriate for a word in a valley, where there is no pitch prominence. The main stress becomes a deaccented—minimized—vowel in a valley.

The citation form of a dictionary word does nothing to help students with the **level** of the main stress in the two-peak profile. Finding only the position of main stress in a word is the weakness of the old goal for word-stress work. A more adequate goal with respect to word stress must also include giving students practice modifying the level of stress in a word to match its place in the two-peak profile.

NEW goal: Teach the position of word stress and the *flexibility to adjust the level of word stress* at this position according to the role the word plays in rhythm.

The limitations of the old goal reflect the limitations of the old model of rhythm in which every content word in a phrase receives a heavy stress and no content word would be found entirely in a valley. In the actual rhythm of spontaneous speech, where there are usually only two maximized vowels in a phrase, the stress of all other content words is pervasively minimized. Reality at the phrase level is that the stress of any word must be flexible enough to fit anywhere in the two-peak profile. So how do we prepare learners to be flexible with word stress?

First we teach learners how to predict the peaks (and valleys) of the two-peak profile. That is because the rhythm profile is the template that tells learners the level of stress to use for all multisyllabic words in a phrase. Then comes practice that targets the skill of adjusting stress levels appropriately. When the prosody of vowels was the target, our first step in practice was to hold the vowel sound constant and vary the prosody of this one vowel. Now that word stress is the target, we hold the word constant and vary the location of this one word in the two-peak profile.

One exercise type from *Speechcraft* puts the same word in two different positions. The target word is *good.* Students fill in the blanks to describe themselves. Then they read the three lines. Their practice contrasts a maximized vowel in *good* (anchor) with a minimized vowel in *good* (valley).

> *Describe yourself by filling in the blanks with some of the words below. Then read the three lines aloud. Focus on contrasting the stress of the word* good.
>
> I'm good at __chitchat__. But I'm not so good at __listening__, | and I don't want to try.
>
> I'm good at __languages__. But I'm not so good at __math__, | and I don't want to try.
>
> I'm good at ________. But I'm not so good at ________, | and I don't want to try.

math	science	sports	research	writing
organizing	music	chitchat	drawing	listening
languages	soccer	dancing	cooking	teaching

Another exercise type, called "Adjusting Stresses to Match the Two-Peak Profile," puts the same words in all three possible positions of the two-peak profile to practice three levels of stress—for the anchor peak, the primary peak, and a valley. After identifying the target words and constructions, learners practice adjusting their stresses until they can do so smoothly. Here an instructor is cautioning a class about getting the most from their literature review.

> *Highlight the words and constructions that are repeated with three levels of main stress. Listen to the dialogue and practice it until you can easily adjust the level of stress in the flow of the discourse.*
>
> **Situation:** An instructor's caution
>
> Here's a fact about footnotes: When reading research, | footnotes don't seem important |
>
> because they are footnotes. But don't ignore them. Sometimes one that doesn't appear
>
> important | contains a fact | that's important to your work, | a fact you've been looking for.

The motivation for asking learners to identify words with three levels of stress (anchor, primary, and destressed valley) is to engage them in something like a treasure hunt. With their understanding of the two-peak profile, the task is not difficult. Yet it draws their attention to the different stress requirements of the rhythm. In this monologue, learners find the words *fact, footnote,* and *important,* each in the three stress positions of the two-peak profile. Exercises of this kind, with supporting audio, develop learners' flexibility to vary the stress of a word in a phrase according its role in the two-peak profile.

Work on adjusting the level of stress is a word-stress topic that connects stress levels directly to the two-peak profile. Stress level, not stress position, signals the start, middle, and end of the focus of a phrase. The contrasting peaks and valleys of the focus must stand out clearly if listeners are going to find the phrase intelligible.

10. Pitch and its Patterns. The role of pitch that is most familiar to ESL/EFL instructors is in pitch patterns or intonation patterns. These patterns are important because they tell us how to interpret what we hear in the focus of a phrase: Is this core meaning a question or a statement? Is it finished or is there more to come? Is it friendly, businesslike, neutral, or insistent? Is it a true question or a request for confirmation? However, intonation patterns are a secondary role of pitch, not its primary role, and therefore not entirely adequate to promote the two-peak profile.

OLD (and incomplete) goal: Teach the pitch of pitch patterns.

The primary role of pitch, along with intensity and duration, is to highlight for listeners the cues that mark the beginning and end of the focus. We recognize these peaks best as listeners, not because speakers maximize the pitch, intensity, and duration of peak vowels, but because they maximize the **difference** between peak vowels and nearby valley vowels. These acoustic markers are the yin and yang (☯) of rhythm.

When it comes to pitch accents, we tend to concentrate more on maximizing the suprasegmentals, including the pitch, of their accented vowels than on minimizing the suprasegmentals of adjacent unaccented vowels. To restore the right balance, we must remember that both ends of the suprasegmental continua deserve equal attention in order to create the difference between these extremes that English listeners expect.

NEW goal: Prioritize the focus first by preserving its yin-yang balance: Teach peaks by contrasting them with valleys. Then teach pitch patterns to help listeners interpret the focus.

If peaks do not stand out, what is the point of an intonation pattern to a listener who does not know what the core message is?

CONCLUSION

This paper began with the assumption that the purpose of oral language is to communicate meaning. If this assumption is accurate, then how does oral language advance this purpose? For English, the answer seems to be that the peaks and valleys of its phrase rhythm identify for listeners the basic pragmatic structure called the focus of the phrase, its core meaning.

This answer says, in effect, that phrase rhythm is the most fundamental linguistic pattern in oral English because of its centrality to intelligible communication. Furthermore, in one way or another, the rest of phonology works to create and build on this rhythm for the ultimate purpose of communicating meaning. That is, all of phonology, including rhythm, is driving toward the goal of making intelligible oral communication possible.

Arriving at these conclusions was the result of taking a fresh look at my assumptions. The fact is that, like many ESL/EFL instructors, I learned my stress-timed-rhythm lessons so well that they created blind spots in my understanding of the communicative role of rhythm.

However, the real turning point for me was the exercise of trying to teach the two-peak profile to my ESL students. It forced me to reexamine everything I was saying about pronunciation, to identify implications of this rhythmic profile for my teaching, and to start implementing changes in content and practice that I saw were needed for the sake of my students.

The critical changes involve having students review parts of speech to facilitate their use of accent rules, starting the course with the two-peak profile, keeping in front of students the centrality of this profile for communicating the semantic essence of a phrase, emphasizing that the defining features of spontaneous speech—short phrases, only one or two peaks, and compressed valleys—help their listeners grasp the semantic essence quickly, and training students in flexibly creating peaks and valleys by manipulating the suprasegmentals of vowels.

The efforts that my pronunciation team and I made to introduce these changes across our entire syllabus have been rewarded by the progress our students have shown. Since they know what they are doing and why, they seem more determined to change their oral skills in the ways we recommend. Our impression, yet to be confirmed empirically, is that they are improving their intelligibility much more quickly now than in the era before the two-peak profile and before being inundated by the myriad ripples of this model of rhythm.

ABOUT THE AUTHOR

Wayne Dickerson is professor emeritus in the Department of Linguistics at the University of Illinois at Urbana-Champaign where he directed the MA-TESOL program and taught courses in pedagogical English phonology (online and on campus) for TESOL candidates and ESL pronunciation for university-level students. He has since helped to develop an online pedagogical phonology course for candidates in the MA-TESOL program at Wheaton College. His research focuses on pedagogical applications of phonetics and phonology, pronunciation pedagogy, the value of orthography for learners, phonological variability, and pronunciation assessment. His two pronunciation textbooks are *Stress in the Speech Stream: The Rhythm of Spoken English* (1989/2004), and associated audio recordings, University of Illinois Press, and (with co-author Laura D. Hahn) *Speechcraft: Discourse Pronunciation for Advanced Learners* (1999), and associated workbooks and audio recordings, University of Michigan Press.

Contact information:
7 Hale Haven Court
Savoy, IL 61874 U.S.A.
E-Mail: dickrson@illinois.edu

REFERENCES

Arvaniti, A. (2012). The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics, 40*, 351-363.

Bolinger, D. (1961). Three analogies. *Hispanica, 44,* 135-136.

Bolinger, D. (1986). *Intonation and its parts.* Stanford, CA: Stanford University Press.

Brazil, D. (1994). *Pronunciation for advanced learners of English. Student's book.* Cambridge: Cambridge University Press.

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

Cauldwell, R. (2002). The functional irrhythmicality of spontaneous speech: A discourse view of speech rhythms. *Apples 2*, 1-24.

Cauldwell, R. (2013). *Phonology for listening, teaching the stream of speech.* Birmingham, UK: Speech in Action.

Dale, P. & Poms, L. (1994). *English pronunciation for international students.* Englewood Cliffs, NJ: Prentice-Hall Regents.

Dauer, R. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics, 11,* 51-62.

Dickerson, W. (2004). *Stress in the speech stream, the rhythm of spoken English.* Urbana, IL: University of Illinois Press. [CD-ROM] (Original work published 1989).

Dickerson, W. (2015). A NAIL in the coffin of stress-timed rhythm. In J. Levis, R. Mohamed, M. Qian & Z. Zhou (Eds.), *Proceedings of the 6th Pronunciation in Second Language Learning and Teaching Conference* (ISSN 2380-9566), Santa Barbara, CA (pp. 184-196). Ames, IA: Iowa State University. https://apling.engl.iastate.edu/alt-content/uploads/2015/05/PSLLT_6th_Proceedings_2014.pdf

Dickerson, W. & Hahn, L. (in press). *Speechcraft: Discourse pronunciation for academic communication* (2nd ed.). Ann Arbor: University of Michigan Press.

Fries, C. (1945). *Teaching & learning English as a foreign language.* Ann Arbor: University of Michigan Press.

Gilbert, J. (2012). *Clear speech: Pronunciation and listening comprehension in North American English* (4th ed.). Cambridge: Cambridge University Press.

Gleason, Jr., H. (1961). *An introduction to descriptive linguistics* (rev. ed.) New York: Holt, Rinehart and Winston.

Gorsuch, G., Meyers, C., Pickering, L., & Griffee, D. (2013). *English communication for international teaching assistants* (2nd ed.). Long Grove: IL: Waveland Press.

Grant, L. (2017). *Pronunciation for clear communication* (4th ed.). Boston: National Geographic Cengage.

Hahn, L. & Dickerson, W. (1999). *Speechcraft: Discourse pronunciation for advanced learners*. Ann Arbor: University of Michigan Press.

Kingdon, R. (1958). *The groundwork of English intonation.* London: Longmans, Green and Co.

Kjellin, O. (1999). Accent addition: Prosody and perception facilitate second language learning. In O. Fujimura, B.D. Joseph, & B. Palek (Eds.), *Proceedings of LP '98 (Linguistics and Phonetics Conference)* at Ohio State University, Columbus, OH, September 1998 (pp. 373-398). Prague: The Karolinum Press. http://olle-kjellin.com/SpeechDoctor/ProcLP98.html.

Ladd, D. (1996). *Intonational phonology.* Cambridge: Cambridge University Press.

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

Miller, S. (2006). *Targeting pronunciation, communicating clearly in English* (2nd ed.). Boston: Houghton Mifflin Company.

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

Pickering, L. (2018). *Discourse intonation: A discourse-pragmatic approach to teaching the pronunciation of English.* Ann Arbor: University of Michigan Press.

Pike, K. (1945). *The intonation of American English.* Ann Arbor: University of Michigan Press.

Prator, C., Jr. (1951). *Manual of American English pronunciation for adult foreign students.* Berkeley, CA: University of California Press.

Roach, P. (1982). On the distinction between 'stress-timed' and 'syllable-timed' languages. In D. Crystal (Ed.), *Linguistic Controversies* (pp. 73-79). London: Edward Arnold.

Tench, P. (1996). *The intonation systems of English.* London: Cassell.

Wells, J. (2006). *English intonation: An introduction*. Cambridge: Cambridge University Press.

---

[i] Note that all of the examples above contrast peak and valley versions of full vowels. The other contrast, not illustrated, is between a peak vowel and an unstressed vowel. Peak-valley switches—*tráffĭc, wálkĭng,* and *brúisĕs*—and valley-peak switches—*Măría, cŏmpóse,* and *cŏllíde*—are also practiced in vowel lessons.

Hayes-Harb, R., & Barrios, S. (2019). Investigating the phonological content of learners' "fuzzy" lexical representation for new L2 words. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching conference*, ISSN 2380-9566, Iowa State University, September 2018 (pp. 55-69). Ames, IA: Iowa State University.

INVITED TALK

INVESTIGATING THE PHONOLOGICAL CONTENT OF LEARNERS' "FUZZY" LEXICAL REPRESENTATION FOR NEW L2 WORDS

Rachel Hayes-Harb, University of Utah
Shannon Barrios, University of Utah

Adult learners are known to experience difficulty using novel second language (L2) phonological contrasts to distinguish words. Indeed, even the ability to perceive and produce a novel contrast with relative accuracy does not guarantee an ability to implement the contrast to distinguish words in tasks requiring lexical access. These observations lead to questions regarding the phonological content of learners' lexical representations of difficult L2 contrasts. In the present study we use an artificial lexicon design involving naïve L2 learners to examine the lexical encoding and retrieval of words containing novel L2 phonological contrasts. In each of two experiments, native English speakers were taught a set of six Japanese-like auditory minimal pairs along with pictured meanings. The members of each pair were differentiated by either consonant length (e.g., [teki] vs. [tekki]) or vowel length (e.g., [teki] vs. [teeki]), which are contrastive in Japanese but not in English. Participants were then tested on their ability to match the pictures to auditory words. These test items were matched (e.g., see picture of 'teki', hear [teki]) or mismatched (e.g., see 'teki', hear [teeki] for segment length). The results are compared to the predictions of several possible scenarios with respect to the ways in which participants might encode and retrieve words varying in segment length.

INTRODUCTION

It is well known that second language (L2) learners exhibit difficulty learning and processing L2 words' phonological forms, in particular when the words involve 'new' phonemes (e.g., Amengual, 2016; Broersma, 2012; Escudero, Hayes-Harb, & Mitterer, 2008; Hayes-Harb & Masuda, 2008; Pallier, Colome, & Sebastian-Galles, 2001; Sebastián-Gallés, Echeverría, & Bosch, 2005; Weber & Cutler, 2004). Several studies have suggested that this difficulty may arise due to the activation of inappropriate lexical items when listening to L2 speech, resulting in slow and/or inaccurate word recognition (e.g., Barrios, Jiang, & Idsardi, 2016; Broersma & Cutler, 2008, 2011; Cutler, Weber, & Otake, 2006; Darcy, Daidone, & Kojima, 2013; Ota, Hartsuiker, & Haywood, 2009; Pallier et al., 2001; Sebastián-Gallés et al., 2005; Sebastian-Gallés, Rodríguez-Fornells, De Diego-Balaguer, & Díaz, 2006). For example, Pallier et al. (2001) demonstrated that highly-proficient Spanish-dominant bilinguals exhibit repetition priming for Catalan minimal pairs differing in a vowel contrast that occurs in Catalan but not in Spanish (e.g., /sol/ 'sun' – /sɔl/ 'ground'), while Catalan-dominant bilinguals do not. Sebastián-Gallés et al. (2005) and Sebastian-Gallés et al. (2006) similarly showed that Spanish dominant bilinguals have difficulty rejecting nonwords that differ from real Catalan words in Catalan-specific vowel contrasts in a lexical decision task, and Sebastian-Gallés et al. (2006) further demonstrated that these erroneous lexical decisions do not elicit error-related negativity as detected by ERP.

The activation of inappropriate lexical competitors during auditory word processing has been interpreted as evidence that learners have non-target-like perceptual and/or lexical representations for difficult novel contrasts. In the case of non-target-like perceptual representations, learners may perceptually neutralize a difficult auditory contrast in favor of the member of the contrast that is most similar to the native category. This perceptual difficulty could play out in at least two different ways, depending on the status of the contrast in the lexicon. First, a neutralized perceptual representation will equally access neutralized lexical representations (i.e., minimal pairs are encoded as homophones; Escudero et al., 2008; Llompart & Reinisch, 2017). Alternatively, a neutralized perceptual representation might cause learners to access the word containing the native ('old') category in cases where the contrast is encoded lexically. Evidence for this latter possibility has been provided by eye-tracking studies demonstrating that learners initially activate words containing the 'old' category even when auditory inputs contain the 'new' category *but not the reverse* (Cutler et al., 2006; Weber & Cutler, 2004), and appears to arise when learners have access to evidence for the contrast in the form of written (Escudero et al., 2008) or articulatory information (Cutler et al., 2006; Escudero et al., 2008; Llompart & Reinisch, 2017; Weber & Cutler, 2004).

On the other hand, learners may have distinct perceptual representations but neutralized or phonologically ambiguous lexical representations for novel contrasts. Non-target-like L2 lexical encoding is in focus in the present study; in particular cases where learners experience lexical confusion due to lexical representations that are contrastive but "fuzzy" (e.g., Cook & Gor, 2015; Darcy et al., 2013; Hayes-Harb & Masuda, 2008).

Non-native-like lexical encoding and fuzzy L2 lexical representations

Among the challenges L2 learners face is navigating the perceptual and lexical-phonological consequences of novel contrasts. In some cases, learners appear to be able to perceptually distinguish novel contrasts but nonetheless experience difficulty in auditory word recognition tasks. Evidence for such cases has come from studies revealing asymmetric response patterns in lexical tasks (e.g., Barrios et al., 2016; Cutler et al., 2006; Darcy et al., 2013; Hayes-Harb & Masuda, 2008; Llompart & Reinisch, 2017, 2018; Sebastián-Gallés et al., 2005; Sebastian-Gallés et al., 2006; Weber & Cutler, 2004).

In one such study, Hayes-Harb and Masuda (2008) investigated native English speakers' memory for Japanese-like words contrasted by singleton and geminate medial consonants. Japanese contrasts singleton and geminate consonants (e.g., /haken/ *dispatch* vs. /hakken/ *discovery*) and short and long vowels (e.g., /kado/ *corner* vs. /kaado/ *card*). These length contrasts are often cited as a source of difficulty for native speakers of English, for whom consonant length is not phonemic (though long consonants can occur across word boundaries, as in *topic* vs. *top pick*), and for whom vowel duration covaries with vowel quality (e.g., /ɛ/ vs. /eɪ/) and vowel lengthening occurs allophonically (i.e., preceding voiced consonants as in *bad* vs. *bat*). Indeed, it has been demonstrated that Japanese length contrasts are difficult for native English speakers to produce (e.g., Han, 1992; Hardison & Saigo, 2010; Hayes-Harb & Masuda, 2008). On the other hand, Japanese consonant and vowel length contrasts may be relatively easy for native English speakers to perceive (see, e.g., Darcy et al., 2013; Tajima, Kato, Rothwell, Akahane-Yamada, & Munhall, 2008; Tajima, Kato, Rothwell, & Munhall, 2003; Tsukada, Cox, Hajek, & Hirata, 2018).

Hayes-Harb and Masuda (2008) taught native English speakers who had no prior Japanese language study, native English learners with one year of Japanese language study, and native Japanese speakers a set of four singleton-geminate minimal pairs (e.g., [pete] and [pette]; the target words) plus four filler words by pairing pictured meanings with auditory forms. They then examined the participants' memory for the words' phonological forms via an auditory word – picture matching task and a picture naming task. In the listening task, detectability (dprime) of the difference between matched (e.g., picture of 'pete' paired with auditory [pete]) and mismatched (e.g., 'pete' – [pette]) items was 1.01 for the participants with no Japanese language experience, 2.11 for the learners of Japanese, and 2.73 for the native Japanese speakers. A new analysis of this data reveals that the dprimes for all three participant groups are significantly greater than zero ($t(11)<.005$ for all). However, on the picture naming task, participants with no Japanese experience produced target geminate consonants only 17% of the time (in contrast to 99% and 58% geminate production by the native Japanese and the learners of Japanese, respectively). Hayes-Harb and Masuda (2008) interpreted the discrepancy between performance on the listening and production tasks by the participants with no Japanese language experience as evidence for partial encoding of the phonological contrast in their memory for the words. They hypothesized this pattern of performance could be accounted for by encoding 'old' phonemes accurately (e.g., [t] as /t/ in /pete/) but the 'new' phonemes as differing from the 'old' phonemes in a phonologically ambiguous, or fuzzy, way (e.g., [tt] as /t*/ in /pette/). Listening task accuracy thus would require only that participants detect whether the input contained /t/ or /t*/, while production accuracy would require a phonologically specific representation, i.e., one that identified length as the distinguishing feature.

Darcy et al. (2013) investigated the lexical encoding of Japanese consonant length contrasts by native Japanese speakers and native English speakers at intermediate and advanced levels of Japanese language proficiency. Darcy et al. (2013) used a lexical decision task involving real Japanese words containing singleton "old" and geminate "new" consonants, and nonwords created by changing a medial consonant from singleton to geminate and vice-versa. They hypothesized that if participants perceived consonant length in the auditory input but their lexical representations for words containing geminate consonants were fuzzy, they should exhibit the following ordinal accuracy (from highest to lowest): (1) Real singleton words (e.g., [akeru] 'to open') should be easy to accept because the input matches the lexical representation. (2) Real geminate words (e.g., [kippu] 'ticket') should also be easy to accept because while they do not exactly match the fuzzy underlying representation, they do not mismatch, either. (3) Geminate nonwords created from real singleton words (e.g., *[akkeru]) should be relatively easy to reject because they contain the new category but are compared to a lexical representation that contains the old category. (4) Singleton nonwords (e.g., *[kipu]) should be difficult to reject because they must be compared to fuzzy lexical representations. They found exactly this pattern in the results for the two learner groups, and concluded that learners experience difficulty with Japanese consonant length contrasts at the lexical level, proposing the convention "?" for indicating fuzziness (an alternative to the '*' proposed by Hayes-Harb & Masuda, 2008).

The Hayes-Harb and Masuda (2008) and Darcy et al. (2013) studies thus provide two different types of evidence for a phonologically ambiguous, or fuzzy, lexical representation of Japanese consonant length by native English speakers: in the former, the discrepancy between performance on an auditory word-picture matching task and a picture naming task, and in the latter, asymmetries in lexical decision performance. A question that emerges, then, is *What does it mean for a lexical representation to be fuzzy?* Thus far "fuzzy" has been taken to mean that a new phonological element is an imperfect match to a native (old) element. In the case of phonological length encoding by native English speakers, one possibility is that fuzziness is associated with a single segment, where, e.g., singleton /t/ and geminate /tt/ are represented by learners as /t/ and /t*/, respectively. However, a number of other scenarios are possible; these are laid out in detail below.

METHODS

This study involved a series of two experiments employing the artificial lexicon paradigm, examining the acquisition of Japanese-like length contrasts by native speakers of English. The studies discussed above investigated the acquisition of consonant length; in the present research we consider the acquisition of both consonant and vowel length in order to determine whether learners perform similarly on both types of phonological length. The first experiment focused on native English speakers' ability to learn and process Japanese-like words contrasted by consonant length; the second examined their ability to learn and process words contrasted by vowel length. Each experiment is presented in turn.

Consonant length experiment

Participants. Twenty-three native speakers of English (13 female, 10 male, ages 18-44, mean age of 22) were recruited from the University of Utah community. Participants did not report any hearing, speech, or language disorders, and none reported taking medications that may affect their cognitive or motor functions. Three reported also speaking Spanish as a native language and one Greek. They reported having studied Spanish (n=13), French (n=3), and ASL (n=1), Portuguese (n=1), and Mandarin (n=1), but had not studied Japanese or any other language with vowel and/or consonant length contrasts (e.g., Arabic, Hindi, or Italian).

Materials. A phonetically-trained female native speaker of Japanese produced six CVCV minimal pairs exemplifying the singleton-geminate contrast in medial position. Most but not all of the words are nonwords in Japanese. Each word was randomly assigned a meaning represented by a line drawing. Minimal pairs were included to increase the likelihood that participants would notice and lexically encode the length contrasts.

Table 1

The six short-long consonant minimal pairs and pictured meanings

Singleton		**Geminate**	
[teki]	*basket*	[tekki]	*book*
[hako]	*car*	[hakko]	*corn*
[hosa]	*flower*	[hossa]	*shirt*
[meso]	*violin*	[messo]	*grapes*
[kite]	*foot*	[kitte]	*house*
[keto]	*spider*	[ketto]	*stove*

Procedure. The experiment involved three phases. In the first, the word learning phase, participants were exposed to the auditory words and their pictured meanings. Each auditory word-picture pair presented once per block, and the block was presented eight times, in a new random order each time. No response was required of participants, who were told only to "learn the words and their meanings as well as possible". The criterion test phase immediately followed word learning. It involved a two-way forced-choice auditory word-picture matching task where each of the twelve words was presented twice, once in each of two conditions: matched and mismatched. In matched trials, the correspondence between the auditory word and the picture were in accordance with the word learning phase, and in mismatched trials, the auditory word was paired with a picture that corresponded to a different word during word learning (see Table 2). As the purpose of the criterion test phase was to ensure that participants had generally learned the words before moving on to the final test, items tested only participants' ability to discriminate among very different words, not the minimal pairs (which were in focus in the final test phase). Participants repeated the word learning and criterion test phases until they reached 90% accuracy on the criterion test.

Table 2

Example criterion test items

Condition	**See**	**Hear**	**Correct Response**
Match		[teki]	Yes (matched)
Mismatch	/teki/	[hako]	No (mismatched)

The final test phase also involved an auditory word-picture matching task. Each picture was presented three times, once in each of three conditions (see Table 3). In the matched condition, the auditory word correctly matched the picture. In the mismatched consonant condition, the auditory word was the geminate counterpart of the correct (singleton) word. In the mismatched vowel condition, the auditory word had the correct consonant length, but the vowel in the first syllable was lengthened. In this way, there was a total of 36 final test trials.

Table 3

Example final test items

Condition	See	Hear	Correct Response
Match		[teki]	Yes (matched)
Mismatch C Length		[tekki]	No (mismatched)
Mismatch V Length	/teki/	[teeki]	No (mismatched)
Match		[tekki]	Yes (matched)
Mismatch C Length		[teki]	No (mismatched)
Mismatch V Length	/tekki/	[teekki]	No (mismatched)

Results. Before presenting the experimental results, we will lay out the predictions of several possible scenarios with respect to the lexical encoding and processing of Japanese consonant length contrasts by native English speakers. The first is the <u>target-like scenario</u> (see Table 4), where participants perceive and encode the novel length contrasts in a target-like way. This should result in high accuracy in all item conditions.

Table 4

Accuracy predictions for the target-like scenario

	Lexical Rep.	Match	Mismatch C	Mismatch V
Short C [old]	/teki/	[teki] Easy to accept	[tekki] Easy to reject	[teeki] Easy to reject
Long C [new]	/tekki/	[tekki] Easy to accept	[teki] Easy to reject	[teekki] Easy to reject

It is worth noting that we predict the same pattern of performance if participants simply annotate each lexical entry with the number of moras (see Table 5); for this reason, a finding of high accuracy in all conditions would require follow-up experiments with additional foil conditions in order to determine the appropriate interpretation.

Table 5

Accuracy predictions for the mora-counting scenario

	Lexical Rep.	Match	Mismatch C	Mismatch V
Short C [old]	/$\text{teki}_{\text{2MORAS}}$/	[teki] (2 moras) Easy to accept	[tekki] (3 moras) Easy to reject	[teeki] (3 moras) Easy to reject
Long C [new]	/$\text{teki}_{\text{3MORAS}}$/	[tekki] (3 moras) Easy to accept	[teki] (2 moras) Easy to reject	[teekki] (4 moras) Easy to reject

In the <u>fuzzy segment scenario</u> (see Table 6), participants have correctly associated the contrast with the target segment, but the representation of that segment is fuzzy. This is the scenario reported by Darcy et al. (2013) and Hayes-Harb and Masuda (2008). In this case, participants have encoded /teki/ and /tek?i/. As a result, [teki] for /tek?i/ should be difficult to reject because the auditory input does not mismatch the fuzzy representation of the segment. By contrast, [teeki] and [teekki] for /teki/ and /tek?i/, respectively, should be easy to reject because the long vowel is not accommodated by the lexical representations.

Table 6

Accuracy predictions for the fuzzy segment scenario

	Lexical Rep.	**Match**	**Mismatch C**	**Mismatch V**
Short C [old]	/teki/	[teki] Easy to accept	[tekki] Easy to reject	[teeki] Easy to reject
Long C [new]	/tek?i/	[tekki] Easy to accept	[teki] Difficult to reject	[teekki] Easy to reject

In the <u>word length scenario</u> (see Table 7), participants have correctly identified length as differentiating lexical items, but have not associated length with a particular segment. In this scenario, [tekki] for /teki$_{\text{LONG}}$/ should be easy to accept because of it contains a long segment, and [teki] for /teki$_{\text{LONG}}$/ should be easy to reject because nothing is long. However, [teekki] for /teki$_{\text{LONG}}$/ should be difficult to reject because it has at least one long segment.

Table 7

Accuracy predictions for the word length scenario

	Lex. Rep.	**Match**	**Mismatch C**	**Mismatch V**
Short C [old]	/teki/	[teki] Easy to accept	[tekki] Easy to reject	[teeki] Easy to reject
Long C [new]	/teki$_{\text{LONG}}$/	[tekki] Easy to accept	[teki] Easy to reject	[teekki] Difficult to reject

Finally, in the <u>fuzzy word scenario</u> (see Table 8), participants have encoded only an indication that the word departs from English in some way. In this scenario, for /(teki)?/ participants should accept all auditory forms because the entire representation is fuzzy.

Table 8

Accuracy predictions for the fuzzy word scenario

	Lexical Rep.	Match	Mismatch C	Mismatch V
Short C [old]	/teki/	[teki] Easy to accept	[tekki] Easy to reject	[teeki] Easy to reject
Long C [new]	/(teki)?/	[tekki] Easy to accept	[teki] Difficult to reject	[teekki] Difficult to reject

Other scenarios are possible, including ones where participants only reject auditory forms that were not presented during the exposure phase (Word Familiarity Scenario), or where they only reject auditory forms that contain new phones (English Bias Scenario). These scenarios are presented in the Appendix.

Table 9 presents the mean proportion correct and mean dprimes, and Figure 1 presents a graph of the dprime data. The consonant length experiment results were submitted to an ANOVA with dprime as the dependent variable and word condition (2 levels: Long Consonant [new], Short Consonant [old]) and mismatch condition (Mismatch Consonant, Mismatch Vowel) as between-subjects variables. Word condition was not significant ($F(1,22)=.680$, $p=.418$, partial eta squared =.030), mismatch condition was not significant ($F(1,22)=.050$, $p=.825$, partial eta squared =.002), but the interaction of the two was significant ($F(1,22)=7.745$, $p=.011$, partial eta squared =.260). Planned comparisons between performance on words containing the 'old' and 'new' phones indicate that in the Mismatch C condition, performance was more accurate on words containing the old phone (Short C; $F(1,22)=4.861$, $p=.038$, partial eta squared =.181) but in the Mismatch V condition, performance was significantly more accurate on the words containing the new phone (Long C) ($F(1,22)=5.039$, $p=.035$, partial eta squared =.186).

Table 9

Mean proportion correct and mean dprime by word condition and mismatch condition

n=23	Mean proportion correct (stdev)			Mean dprime (stdev)	
	Matched	Mismatch C	Mismatch V	Mismatch C	Mismatch V
Short C [old]	.891 (.119)	.457 (.373)	.319 (.359)	1.013 (1.031)	.674 (.992)
Long C [new]	.804 (.199)	.399 (.354)	.558 (.375)	.732 (.969)	1.130 (1.037)

Looking first at the Mismatch C results, we see the asymmetry in responses predicted by the fuzzy segment scenario and the fuzzy word scenario, with significantly lower dprimes for Long C [new] words than for Short C [old] words. In addition, the significant difference reported in the Mismatch V results is the *opposite* of that predicted by the fuzzy word scenario. The results of this experiment are thus consistent with participants having encoded the old Short C words using the native category, and the 'new' Long C words in a phonologically ambiguous, or fuzzy, way. The pattern of results further suggests that participants may have correctly associated the fuzziness with the medial consonant, or at least that they do not incorrectly associate the fuzziness with the preceding vowel. Since Japanese has both vowel and consonant length contrasts, a second experiment involving the representation of vowel length allows us to see whether the pattern observed here is also found for words containing vowel, as opposed to consonant, length.

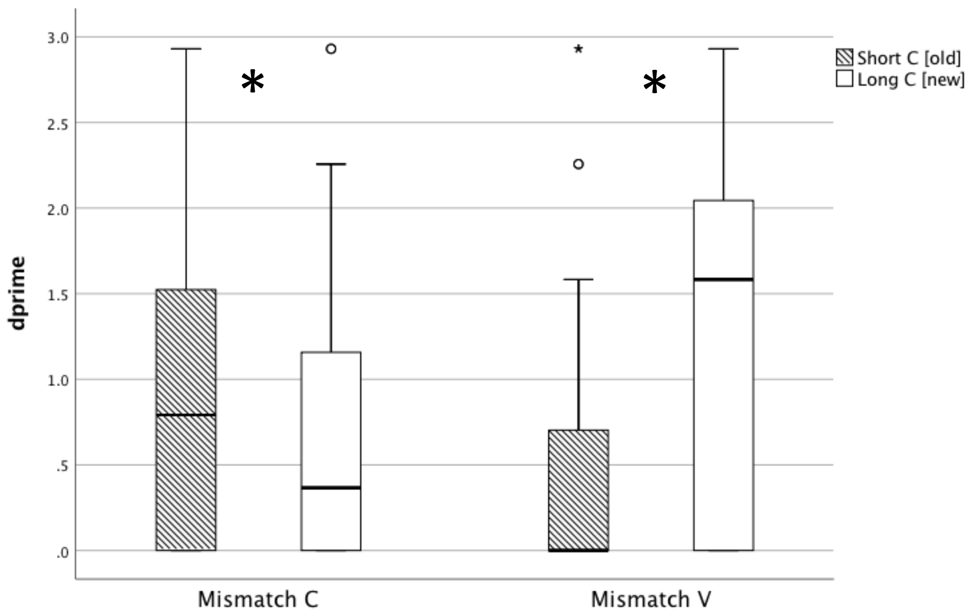


Figure 1. Boxplot of consonant experiment dprimes by pictured word length (Long C [new], Short C [old]) and mismatch condition (Mismatch C, Mismatch V).

Vowel length experiment

Twenty-three new native speakers of English (16 female, 7 male, ages 18-48, mean=22) participated in the vowel length experiment; all met the same criteria as participants in the consonant length experiment. One reported also speaking Spanish as a native language and one Portuguese. They reported having studied Spanish (n=11), French (n=10), ASL (n=1), Cambodian (n=1), Portuguese (n=1), and Chinese (n=1). The materials were six CVCV minimal pairs exemplifying the Japanese short-long vowel contrast in the first syllable: [teki]-[teeki], [hako]-[haako], [hosa]-[hoosa], [meso]-[meeso], [kite]-[kiite], and [keto]-[keeto] and were assigned meanings represented by line drawing pictures. The procedure for the vowel length experiment was identical to that of the consonant length experiment.

Results. Mean proportion correct was converted to dprime (see Table 10).

Table 10

Mean proportion correct and mean dprime by word condition and mismatch condition in the vowel length experiment

n=23	Mean proportion correct (stdev)			Mean dprime (stdev)	
	Matched	Mismatch C	Mismatch V	Mismatch C	Mismatch V
Short Vowel [old]	.855 (.176)	.299 (.265)	.319 (.313)	.547 (.651)	.626 (.796)
Long Vowel [new]	.891 (.139)	.667 (.289)	.312 (.360)	1.538 (.951)	.695 (.848)

The vowel length experiment results were submitted to an ANOVA with dprime as the dependent variable and word condition (2 levels: Long Vowel [new], Short Vowel [old]) and mismatch condition (Mismatch Consonant, Mismatch Vowel) as between-subjects variables. Word condition was significant (F(1,22)=9.933, p=.005, partial eta squared =.311), mismatch type was not significant (F(1,22)=3.349, p=.081, partial eta squared =.132), and the interaction of the two was significant (F(1,22)=13.131, p=.002, partial eta squared =.374). Planned comparisons between performance on words containing the 'old' and 'new' phones indicate that performance was significantly more accurate on words containing the new phone (long vowel) in the consonant mismatch condition (F(1,22)=15.453, p=.001, partial eta squared =.413) but there was no difference between the two word types in the vowel mismatch condition (F(1,22)=.186, p=.671, partial eta squared =.008).

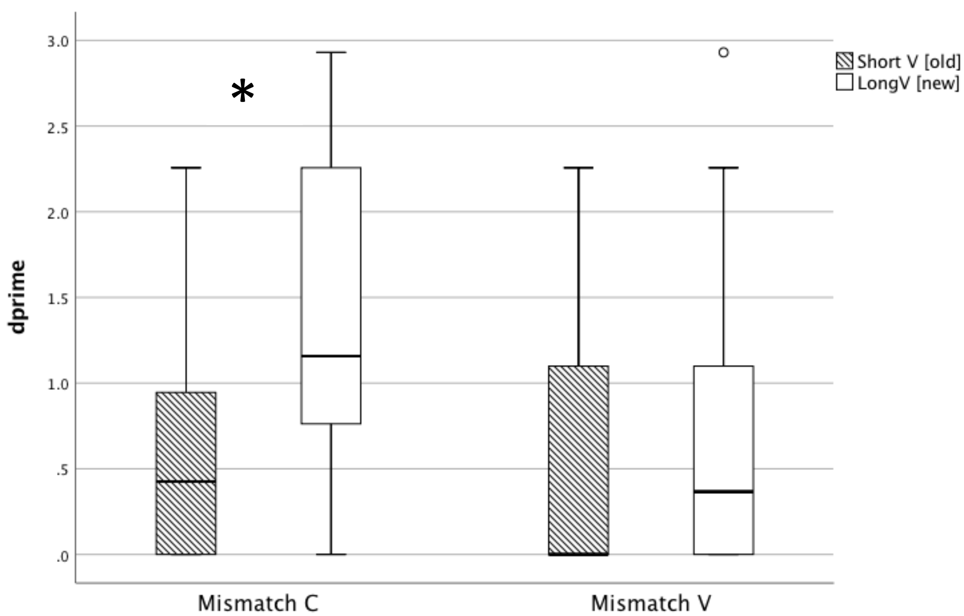


Figure 2. Vowel experiment dprimes by pictured word length (Long Vowel [new], Short Vowel [old]) and mismatch condition (Mismatch Consonant, Mismatch Vowel).

We did not see an asymmetry in responses to the Mismatch V items, and thus do not find evidence in support of either the fuzzy segment scenario or the fuzzy word scenario; indeed, this pattern of results suggests that participants have not encoded short- and long-vowel words contrastively in their lexical representations. Consistent with the consonant experiment, we do observe significantly higher accuracy in the Mismatch C – Long V condition. In both experiments, participants are able to reject the four-mora [teekki] items, suggesting that these items saliently mismatch the word forms learned during training, possibly due to their extra-long duration.

DISCUSSION

In the experiments reported here we sought to investigate the so-called "fuzziness" of L2 lexical representations. We presented a number of possible scenarios with respect to native English speakers' performance on an auditory word – picture matching task in effort to characterize the locality of information that is associated with the L2 fuzzy representation of consonant and vowel length. In the consonant length experiment, where participants learned Japanese-like singleton-geminate minimal pairs, we found evidence for the fuzzy representation of consonant length, evidenced by asymmetric detection of consonant length mismatches associated with words containing short (old) and long (new) consonants. We further found they had not incorrectly associated phonological length with vowels in the word, in fact exhibiting an asymmetry in responses for vowel-mismatched items that was in the opposite direction than would be predicted if they had done so. It is important to note, however, that the relevant dprimes were quite small (<1), indicating that even when participants appeared to have established contrastive lexical representations to some extent, their resulting performance was not highly accurate.

In the vowel experiment, we did not find evidence for the contrastive encoding of vowel length—participants performed similarly on vowel-mismatched items for both long vowel and short vowel words. The discrepancy between the results for consonant length, where we found evidence of fuzziness associated with the representation of long consonants, and vowel length, where we did not find any evidence of differential representation of short and long vowels in participants' lexical representations, suggests unsurprisingly that novel L2 contrasts can differ in the difficulty they pose for learners. In this case, under identical exposure conditions, the native English-speaking participants were better able to lexically encode consonant than vowel length, perhaps due to a differential perceptibility of the two contrast types.

The experiments reported here have provided some insight into the locality of fuzziness, indicating minimally that the fuzzy representation of L2 consonant length is not incorrectly associated with a neighboring vowel. However, there is much more to be investigated with respect to the phonological nature of fuzziness: For segmental contrasts, is fuzziness associated with individual features or with a segment as a whole? What role does "category goodness" (Best, 1995) play in determining learners' assessment of auditory input relative to fuzzy segmental representations? How does the phonological nature of fuzziness evolve over time in learners, and how does fuzziness as conceived here relate to findings elsewhere in the literature concerning "phonolexical robustness" (see, e.g., Llompart & Reinisch, 2018)? The present research, in laying out a number of hypotheses with respect to L2 lexical encoding scenarios, represents a starting point in the investigation of the phonetic/phonological properties of fuzzy lexical representations.

ACKNOWLEDGMENTS

We are grateful to Mizuki Miyashita for producing the study stimuli, members of the Speech Acquisition Lab at the University of Utah for help with data collection and for feedback on the presentation, and to the audience at the 10th annual *Pronunciation in Second Language Learning and Teaching* conference for their feedback.

ABOUT THE AUTHORS

Rachel Hayes-Harb (r.hayes-harb@utah.edu) is a Professor and co-director of the Speech Acquisition Lab in the Department of Linguistics at the University of Utah. Her research is focused on the acquisition of second language sound systems by adult learners, with a particular focus on the effects of various types of auditory and visual input in second language lexical and phonological development.

Shannon Barrios (s.barrios@utah.edu) is an Assistant Professor and co-director of the Speech Acquisition Lab in the Department of Linguistics at the University of Utah. Her research uses a variety of methods to examine the acquisition of the phonology of second languages by adults with the goal of characterizing the representations that guide learners' processing of nonnative input and the mechanisms by which learners' representations change.

REFERENCES

Amengual, M. (2016). Cross-linguistic influence in the bilingual mental lexicon: Evidence of cognate effects in the phonetic production and processing of a vowel contrast. *Frontiers in Psychology, 7*, 617. doi:10.3389/fpsyg.2016.00617

Barrios, S., Jiang, N., & Idsardi, W. J. (2016). Similarity in L2 phonology: Evidence from L1 Spanish late-learners' perception and lexical representation of English vowel contrasts. *Second Language Research, 32*(3), 367-395. doi:10.1177/0267658316630784

Best, C. T. (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). Timonium, MD: York Press.

Broersma, M. (2012). Increased lexical activation and reduced competition in second-language listening. *Language and Cognitive Processes, 27*(7-8), 1205-1224. doi:10.1080/01690965.2012.660170

Broersma, M., & Cutler, A. (2008). Phantom word activation in L2. *System, 36*(1), 22-34. doi:10.1016/j.system.2007.11.003

Broersma, M., & Cutler, A. (2011). Competition dynamics of second-language listening. *Quarterly Journal of Experimental Psychology, 64*(1), 74-95. doi:10.1080/17470218.2010.499174

Cook, S. V., & Gor, K. (2015). Lexical access in L2: Representational deficit or processing constraint? *The Mental Lexicon, 10*(2), 247-270. doi:10.1075/ml.10.2.04coo

Cutler, A., Weber, A., & Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics, 34*(2), 269-284. doi:10.1016/j.wocn.2005.06.002

Darcy, I., Daidone, D., & Kojima, C. (2013). Asymmetric lexical access and fuzzy lexical representations in second language learners. *The Mental Lexicon, 8*(3), 372-420. doi:10.1075/ml.8.3.06dar

Escudero, P., Hayes-Harb, R., & Mitterer, H. (2008). Novel second-language words and asymmetric lexical access. *Journal of Phonetics, 36*(2), 345-360. doi:10.1016/j.wocn.2007.11.002

Han, M. S. (1992). The timing control of geminate and single stop consonants in Japanese: A challenge for nonnative speakers. *Phonetica, 49*(2), 102-127. doi:10.1159/000261906

Hardison, D. M., & Saigo, M. M. (2010). Development of perception of second language Japanese geminates: Role of duration, sonority, and segmentation strategy. *Applied Psycholinguistics, 31*(1), 81-99. doi:10.1017/s0142716409990178

Hayes-Harb, R., & Masuda, K. (2008). Development of the ability to lexically encode novel second language phonemic contrasts. *Second Language Research*, *24*(1), 5-33. doi:10.1177/0267658307082980

Llompart, M., & Reinisch, E. (2017). Articulatory information helps encode lexical contrasts in a second language. *Journal of Experimental Psychology: Human Perception and Performance, 43*(5), 1040-1056. doi:10.1037/xhp0000383

Llompart, M., & Reinisch, E. (2018). Robustness of phonolexical representations relates to phonetic flexibility for difficult second language sound contrasts. *Bilingualism: Language and Cognition*, 1-16. doi:10.1017/s1366728918000925

Ota, M., Hartsuiker, R. J., & Haywood, S. L. (2009). The KEY to the ROCK: Near-homophony in nonnative visual word recognition. *Cognition, 111*(2), 263-269. doi:10.1016/j.cognition.2008.12.007

Pallier, C., Colome, A., & Sebastian-Galles, N. (2001). The influence of native-language phonology on lexical access: Exemplar-Based versus abstract lexical entries. *Psychological Science, 12*(6), 445-449. doi:10.1111/1467-9280.00383

Sebastián-Gallés, N., Echeverría, S., & Bosch, L. (2005). The influence of initial exposure on lexical representation: Comparing early and simultaneous bilinguals. *Journal of Memory and Language, 52*(2), 240-255. doi:10.1016/j.jml.2004.11.001

Sebastian-Gallés, N., Rodríguez-Fornells, A., De Diego-Balaguer, R., & Díaz, B. (2006). First- and second-language phonological representations in the mental lexicon. *Journal of Cognitive Neuroscience*, *18*(8), 1277-1291. doi:10.1162/jocn.2006.18.8.1277

Tajima, K., Kato, H., Rothwell, A., Akahane-Yamada, R., & Munhall, K. G. (2008). Training English listeners to perceive phonemic length contrasts in Japanese. *Journal of the Acoustical Society of America*, *123*(1), 397-413. doi:10.1121/1.2804942

Tajima, K., Kato, H., Rothwell, A., & Munhall, K. G. (2003). *Perception of phonemic length contrasts in Japanese by native and non-native listeners.* Paper presented at the 15th International Congress of Phonetic Sciences, Barcelona, Spain.

Tsukada, K., Cox, F., Hajek, J., & Hirata, Y. (2018). Non-native Japanese learners' perception of consonant length in Japanese and Italian. *Second Language Research*, *34*(2), 179-200. doi:10.1177/0267658317719494

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, *50*(1), 1-25. doi:10.1016/s0749-596x(03)00105-0

APPENDIX

In the English bias scenario, participants do not compare auditory forms heard during the test to lexical representations, but rather only reject auditory forms that are not English-like.

Accuracy predictions for the English bias scenario

	Lexical Rep.	**Match**	**Mismatch C**	**Mismatch V**
Short C [old]	n.a.	[teki] Easy to accept	[tekki] Easy to reject	[teeki] Easy to reject
Long C [new]	n.a.	[tekki] Difficult to accept	[teki] Difficult to reject	[teekki] Easy to reject

In the word familiarity scenario, participants do not compare auditory forms heard during the test to lexical representations, but rather only reject auditory forms that were not presented to them during the word learning phase.

Accuracy predictions for the word familiarity scenario

	Lexical Rep.	**Match**	**Mismatch C**	**Mismatch V**
Short C [old]	n.a.	[teki] Easy to accept	[tekki] Difficult to reject	[teeki] Easy to reject
Long C [new]	n.a.	[tekki] Easy to accept	[teki] Difficult to reject	[teekki] Easy to reject

O'Brien, M. G. (2019). Investigating stress assignment. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Ames, IA, September 2018 (pp. 70-81). Ames, IA: Iowa State University.

RESEARCH WORKSHOP

INVESTIGATING STRESS ASSIGNMENT

Mary Grantham O'Brien, University of Calgary

Correctly emphasizing syllables in words and words in sentences (i.e., producing stress) makes both words and sentences easier to understand. Determining whether L2 learners are able to accurately produce and perceive stress can be difficult, though. This may have to do, among other things, with a researcher's operationalization of stress, data collection procedures, and the ways in which the data are analyzed. This contribution takes researchers through a series of steps for both collection and analysis of L2 learner lexical and sentential stress data.

INTRODUCTION

Research has shown that correctly assigning lexical stress is an important aspect of being understood (e.g., Caspers, 2010; Trofimovich & Isaacs, 2012). If we consider the German example in (1), taken from Kleber and Niebuhr (2010), we can see that the meaning of some words changes depending on which syllable is stressed. When this is the case, we say that stress assignment is contrastive in a language.

(1) August
  a. ˈAugust
  b. Auˈgust

If the word is stressed on the first syllable [ˈ] in German, as in (1a), it is a name. If it is stressed on the second syllable, as in (1b), it is a month. Kleber and Niebuhr investigated the role of context in participants' perception of lexical stress assignment through a forced-choice identification task. Participants had to determine whether they heard (1a) or (1b) when it was presented in a context of another name or another month. When the cues to lexical stress were less robust, participants chose the name *ˈAugust* after they heard the name *Friedrich*. However, in the context of another month, *Juli*, 'July', participants chose the month *Auˈgust*. This study demonstrated the importance of context in the disambiguation of ambiguous lexical stress assignment cues. In the real world, however, the context is often less clear, and research has shown that listeners have difficulty processing speech with lexical stress assignment errors, even if the word being uttered is not a member of a stress minimal pair (i.e., two lexical items that differ only in stress assignment, as in (1), Bond & Small, 1983; van Heuven, 2008).

Lexical and sentential stress

When we speak of lexical stress, we mean the syllable in a word that is emphasized. A number of previous studies have demonstrated the importance of accurate lexical stress assignment in being understood. This is the case for both native (e.g., Field, 2005) and nonnative speakers (e.g., Trofimovich & Isaacs, 2012). Inaccurately assigning lexical stress may lead to slowed lexical access (e.g., van Heuven, 2008) and reduced understanding on the part of the listener (Caspers, 2010). L2 learners’ ability to accurately perceive and produce lexical stress may depend on the pairing of the L1 and the L2. For example, native speakers of French, who do not have contrastive lexical stress assignment in their L1, have demonstrated difficulty perceiving (e.g., Dupoux, Sebastian-Galles, Navarrete, & Peperkamp, 2008) and producing (e.g., Yoon & Heschuk, 2011) variable lexical stress in languages like English.

Sentential stress differs from lexical stress in that it is assigned at the level of the clause. The way in which stress is assigned within a sentence depends on the scope of the focus, which can be broad (i.e., referring to an entire clause) or narrow (i.e., referring to an individual phrase or lexical item, e.g., Ladd, 1980). The focused element is the part of a clause that is emphasized and that can answer an implicit or explicit question (Krifka, 2008). When an utterance is produced out of the blue or in answer to a question like “What’s happening?”, this is considered to be a broad focus, or all-new, utterance. In Germanic languages like German and English, the final content word is emphasized in all-new utterances (Féry, 1993). When speakers produce utterances with narrow focus, they highlight information that is meant to stand out from the rest of the sentence. Narrow focus utterances are often produced in response to questions beginning with question words, as demonstrated in (2). As in (1), stress is indicated through the use of [ˈ] immediately preceding the onset of the stressed syllable.

(2) The children travel every day with the bus to school.
  a. When do the children travel with the bus to school?
     The children travel every ˈday with the bus to school.
  b. Who travels every day with the bus to school?
     The ˈchildren travel every day with the bus to school.
  c. How do the children get to school every day?
     The children travel every day with the ˈbus to school.
  d. Where do the children go every day on the bus?
     The children travel every day with the bus to ˈschool.

Hahn (2004) investigated the role that correct production of sentential stress plays in the understanding of L2 speech. Participants in the study evaluated three versions of a lecture given by a speaker of L2 English: one with correct sentential stress, one with incorrect sentential stress, and one with no sentential stress. Participants both recalled more information from, and they showed a tendency to more easily process, the lecture with correct sentential stress.

EXPERIMENTING WITH STRESS

As is the case with any type of experimental research, it is important to ensure that the data we gather from our participants will enable us to answer our research questions in a meaningful way. When we want to investigate stress assignment, we have a range of options to examine how L2 learners produce and perceive stress.

Designing production experiments

Recent studies have looked at the extent to which L2 learners can assign lexical stress in production. Although it might be possible to make use of pictures to elicit semi-spontaneous utterances containing various target items, most L2 stress assignment experiments utilize a reading task, as this ensures that all of the participants produce the same target items, many of which cannot be illustrated via images. In the experiments, words are often read within a carrier phrase such as "I say the word ___________ again" (e.g., Chen, 2013; Domahs, Plag, & Carroll 2014[i]; Tremblay, 2008). Carrier phrases are used to control for prosodic effects (e.g., rising intonation, contrastive stress).

To date studies looking at L2 learners' production of sentential stress have been somewhat limited. Researchers who have investigated it have used two main types of tasks: contextualized sentence reading (O'Brien & Jackson, 2013) and responses to questions about images (e.g., O'Brien & Féry, 2015; O'Brien & Gut, 2011). In both of these types of tasks participants are provided with a context (i.e., a sentence preceding the reading task in O'Brien & Jackson, 2013 or a question about an image in O'Brien & Gut, 2011) that requires them to highlight a particular word in the utterance being produced. The *Questionnaire on Information Structure* (Skopeteas et al., 2006) provides researchers with materials for a range of studies investigating the production of focus and examining information structure more generally. The materials include a series of images and guidelines for experimental tasks that researchers investigating a variety of languages can use to elicit various types of focus (e.g., given vs. new, all new, contrastive focus).

Designing perceptual experiments

A number of task options are available, and this section presents just three of them: ABX tasks, stress preference perception tasks, and gating tasks. These tasks differ in the extent to which they require participants to rely purely on their discrimination ability when completing the task.

Participants' ability to detect lexical stress has often been examined through the use of ABX tasks or variants thereof (e.g., Correia, Butler, Vigario, & Frota, 2015; Dupoux, Pallier, Sebastian, & Mehler, 1997; Tremblay, 2009). In this task, participants hear three stimuli, A, B, and X, as shown in (3).

(3) Sample ABX task
A: ˈinsert
B: inˈsert
X: inˈsert

The participants are to determine whether the target stimulus, X, is the same as A or B. In this case, the correct answer is B. Researchers have criticized ABX tasks for placing a relatively high processing load on participants, who are required to hold both A and B in short-term memory and compare them both to X, thus resulting in less accurate performance when token A is the same as token X (Tremblay, 2009). Others have indicated that participants in ABX studies do not need to compare both token A and token B to X. Instead, simply determining whether B and X are the same allows participants to complete the task (Beddor & Gottfried, 1995). Researchers have proposed that AXB tasks solve this problem, as this task requires listeners to compare both A and B to X (Strange & Shafer, 2008; Tremblay, 2009).

Other tasks used to determine the extent to which participants perceive lexical stress require learners to rely on a more general lexical stress assignment system (i.e., analogy with known words or stress assignment rules) or some level of lexical encoding. One option available to researchers who are interested in investigating the extent to which learners are able to make use of phonological (e.g., syllable weight) or morphological (i.e., affixes) cues to lexical stress assignment is the stress preference perception task (e.g., Guion, Clark, Harada, & Wayland, 2003; Tight, 2007). In this task participants are presented with a written word as in (4). They then listen to productions thereof that differ in terms of their lexical stress assignment.

(4) outrageous *(presented as a written word)*
a. ˈoutrageous
b. outˈrageous
c. outraˈgeous

When carrying out the task participants are required to listen to each variant of the target item (here a, b, and c) and determine which is correct. Researchers are able to not only measure correctness scores, but they can also measure the number of times participants have listened to each token as evidence of participants’ level of confidence with their choice.

Another option for researchers investigating learners’ abilities to encode cues to lexical stress is the gating task. This task requires participants to listen to increasingly longer portions of a word (often within a sentence) and to determine which of a number of similar words is being produced (Field, 2008; Grosjean, 1983; Grosjean & Hirt, 1996). A German example is provided in (5).

(5) Er sagt Direkˈtoren. (*target sentence, not presented to participants*)
‘He says directors.’

The following three options are provided on the screen:
a. Er sagt Direktor.
‘He says director’
b. Er sagt Direktoren.
‘He says directors.’
c. Er sagt Direktorat.
‘He says directorate.’

Gates presented to participants:
1. Er sagt Di
2. Er sagt Direk
3. Er sagt Direkˈtor

The longest gate ever presented to participants should only be as long as the shortest potential answer. Upon hearing each of the gates, participants are required to make a judgement about which sentence is being produced (a, b, or c) and to provide a confidence rating on a scale from 1-10 (where 1=very unsure and 10=very sure). In example (5) above, the first gate contains no information about lexical stress assignment. We would therefore expect both that listeners would simply guess which sentence (a, b, or c) has been spoken and that they would rate their confidence in their decision very low. By the time participants get to the second gate, which also does not contain any overt cues to lexical stress assignment for the target item, we would expect them to exclude the first item, *Diˈrektor*, which is stressed on the second syllable. We would expect higher confidence ratings, given that participants are able to exclude one of the items. At the third gate, when participants hear the stressed syllable, we would expect that those participants who know that *Direkˈtoren* is stressed on the third syllable would choose option b and provide a high confidence rating. The combination of a participants’ choice along with the confidence rating provides insights into the time course of participants' processing of acoustic cues to lexical stress assignment. Gating tasks are a good fit when dealing with languages that have similar words that differ on the basis of lexical stress assignment (e.g., suffixed words).

Researchers who are interested in making use of a gating task should begin by recording the complete target items. Once it has been determined that the tokens contain both robust acoustic cues to lexical stress assignment and that listeners are able to hear these differences (see the section on acoustic analyses below), the sound files that correspond to the various gates can be prepared. Figures 1 and 2 contain the spectrograms for the first two gates for the sentence *I say po ˈlitical* (which can be distinguished in stress assignment from similar words *ˈpolitics* and *poli ˈtician*).

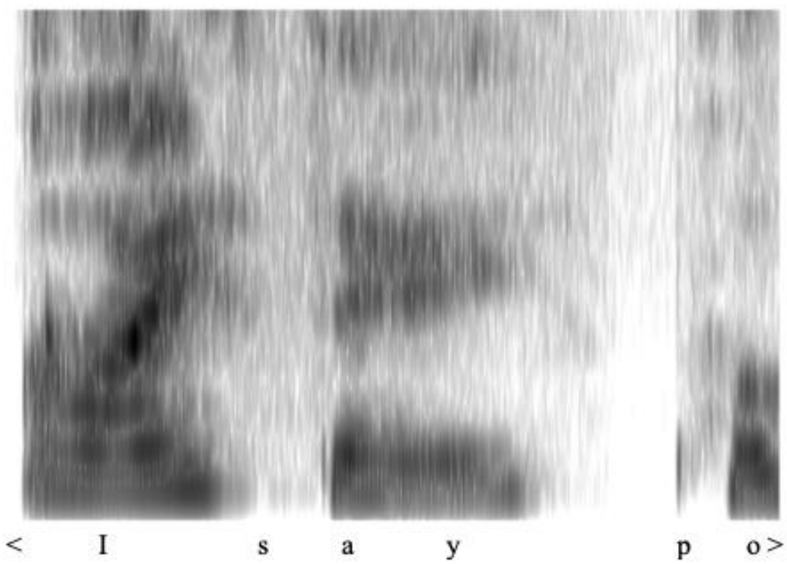


Figure 1. Spectrogram for gate 1 *I say po*.

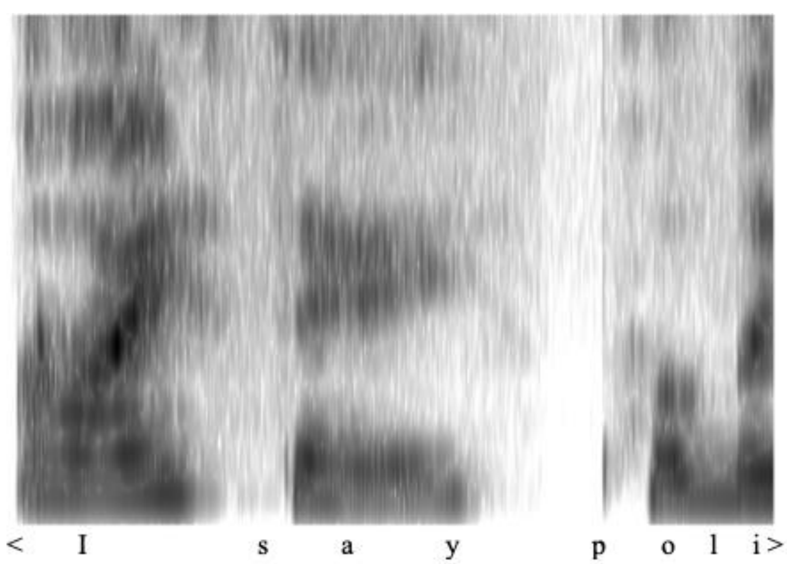


Figure 2. Spectrogram for gate 2 *I say po ˈli*.

When creating gates it is a good idea to save the original files separately in order to ensure that gated files can be recreated in the future if necessary. It is easiest to keep track of the gated files according to the name of the actual target word and the number of syllables that the target word contains (e.g., gate 1 as political1.wav).

Other perception tasks have been utilized in order to gain insights into how participants process lexical stress assignment in an L2. For example, lexical decision tasks and sequence recall tasks provide information about the extent to which L2 learners make use of lexical stress in the encoding of words. In studies investigating lexical stress, lexical decision tasks require participants to determine whether a given stimulus is a word (e.g., Dupoux, Peperkamp, & Sebastian-Galles, 2010). As such, words with incorrect stress assignments should not be categorized as words. In sequence recall tasks, participants are required to recall a series of minimal pairs that they have heard that differ only in lexical stress assignment (Correia et al., 2015; Peperkamp, Vendelin, & Dupoux, 2010). Participants in Correeia et al. (2015) were required, after hearing a series like [ˈnumi]-[nuˈmi]-[nuˈmi]-[ˈnumi]-[nuˈmi], to remember the order in which they heard the words and reproduce the order. The authors note that the task places a heavy processing load on participants.

Choosing tokens

When deciding on which tokens to use in lexical stress assignment studies, researchers should consider a number of important factors. These include a word's frequency, participants' familiarity with the word in question, and the cognate and real word status thereof. While a word's frequency as determined by referring to a spoken or written corpus may play a role in L2 learners' perception or production of lexical stress, familiarity may be a more important factor for L2 learners. For example, the most frequent words that classroom language learners are exposed to often differ from those that are most frequent in a corpus. One possibility for determining participants' familiarity with tokens is to have them rate familiarity at the end of an experiment. A rating scale might range from 1 ("I have never seen this word. I do not know its meaning or how to use it.") to 5 ("I know the word, its meaning, and how to use it."), as in Maczuga, O'Brien, and Knaus (2017).

Research has demonstrated that cognates have a special status in the L2, and participants' ability to correctly assign stress to cognates may differ from their assignment to non-cognates (e.g., Lord, 2001; Maczuga et al., 2017). Thus, if researchers decide to make use of cognates in their studies, they should control for and test the effects of a word's cognate status, for example, by determining if participants perform differently on cognate words as opposed to non-cognate words. One final factor to consider is whether to make use of nonsense words. Whereas tasks using real word tokens may tap into participants' lexical encoding abilities, those using nonsense words are frequently used to control for a word's frequency and participants' familiarity with it (e.g., Domahs, 2014; Jarmulowicz, Taran, & Hay, 2008; Tight, 2007).

ANALYZING STRESS

Researchers need to analyze stress assignment in stimuli used in perceptual studies and in the tokens produced in production studies. When preparing stimuli for perceptual studies and when analyzing stress assignment accuracy in production studies, researchers should ensure that tokens contain robust cues to lexical stress assignment. There are two options for analyzing stress assignment accuracy: acoustic analyses and listener judgments.

Acoustic analyses

When performing acoustic analyses, it is important for researchers to determine which acoustic cues are relevant for stress assignment in the given target language. Many languages rely on cues that include duration (i.e., stressed syllables usually take longer to pronounce), intensity (i.e., stressed syllables tend to be louder), and/or pitch (i.e., there is often a change in fundamental frequency in stressed syllables). Acoustic analyses can be carried out in software including *Audacity* (Audacity Team, 2018) or *Praat* (Boersma & Weenink, 2018). Figures 3 and 4 are waveforms, spectrograms, and TextGrids from *Praat* of the German word *Dominos* 'dominos' produced by the same speaker with stress produced on the first and second syllables, respectively. Syllable duration is marked on the syllables tier (S1, S2, S3), and vowel duration is marked on the vowels tier (V1, V2, V3). The yellow line represents intensity, and the blue line represents pitch. Formants are represented via the red dots.

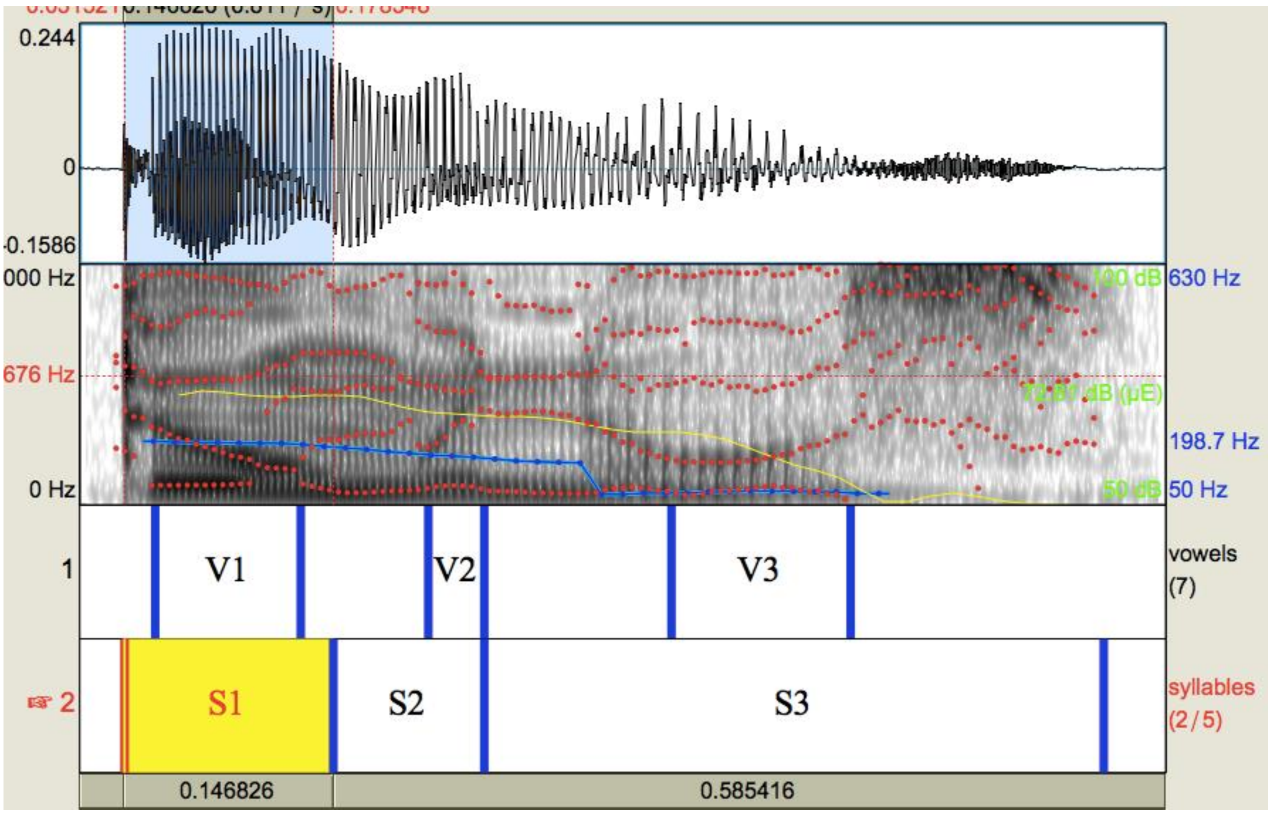


Figure 3. German ˈ*Dominos*.

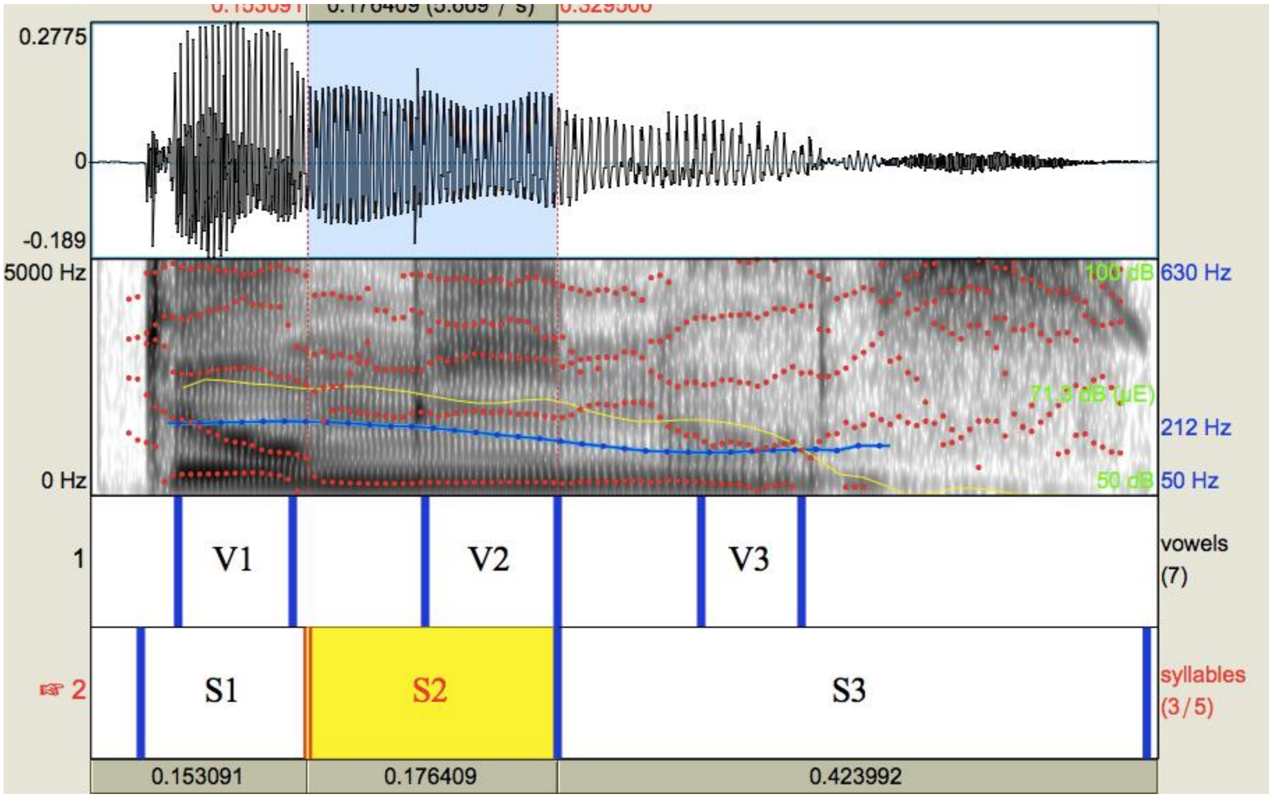


Figure 4. German *Doˈminos*.

Speech segmentation is not a completely straightforward task. For example, when it comes to syllabification of words, it is important for researchers to have a plan for how to handle issues like measuring duration in vowels that have a relatively short steady state and where to divide word-medial consonant clusters. In addition, it is sometimes difficult to determine precisely where one segment ends and the next begins. It is therefore important for researchers to begin segmentation with a clear plan for how to handle issues such as these and for a second researcher to check segmentation. While it is possible to carry out acoustic measurements completely by hand[ii], it is highly recommended that researchers carry out acoustic analyses by creating TextGrids for each target item that they can then submit to batch analyses via a script in *Praat*. Although providing instructions for the creation of TextGrids and the use of scripts is beyond the scope of the current contribution, researchers who wish to make use of these are encouraged to visit the active, highly collaborative *Praat* users’ group.[iii]

Listener judgments

Researchers who are interested in the extent to which stress assignment is perceived in the real world are encouraged to make use of listener judgments, which are considered the gold standard (Derwing & Munro, 2009). In fact, the researchers go so far as to state that “listeners’ judgments are the only meaningful window into accentedness and comprehensibility… [W]hat listeners perceive is ultimately what matters most” (Derwing & Munro, 2009, p. 478).

Determining the presence or absence of stress on a given syllable is not always a clear-cut issue. For example, L2 learners may produce pauses between syllables, thereby making it difficult to determine which syllable is accented. Even if they do produce complete words, learners may produce equal stress on all of the syllables in a word and/or on multiple words in a sentence. When we carry out listener analyses of stress assignment, multiple listeners are often relied upon to determine which syllable is stressed. If the ultimate goal is to determine which syllable speakers have stressed, researchers often rely on the judgments of two listeners (e.g., Maczuga et al., 2017; Yu, 2008). The listeners usually carry out the judgments independently and discuss any disagreements. In the case of a disagreement, the researcher should keep track of such anomalies and have a plan for how to account for these tokens, even if it means removing them from the final analyses.

DISCUSSION

Stress assignment, like other aspects of prosody, is complex. As such, there are fewer studies investigating it than there are studies investigating the perception and production of segments. Nonetheless, given its importance for overall understanding, researchers are encouraged to investigate it in a range of L2s. Moreover, given that lexical stress varies across L1s, it is important to look at a range of L1-L2 pairings. Such studies will provide insights into notions like stress deafness and the role of awareness in stress assignment accuracy. Because the results of studies often depend on the methodology employed, researchers are encouraged to make use of a range of tasks—both perception and production—in order to gain a more nuanced understanding of L2 stress assignment.

ACKNOWLEDGMENTS

I would like to acknowledge research assistants Ross Sundberg, Stéphanie Chicoine, Lisa Süßenbach, and Helene Bramwell for assistance with this work. This research was supported by the Social Sciences and Humanities Research Council of Canada.

ABOUT THE AUTHOR

Mary Grantham O'Brien is a professor of German at the University of Calgary. Her research centers on L2 perception and production. In recent studies she has investigated listener judgments of L2 speech as well as the extent to which L2 learners make use of target language-relevant acoustic cues in their productions. She has carried out a number of perceptual studies to determine the extent to which L2 listeners perceive prosodic cues. Her current research focuses on lexical stress assignment among L2 learners of German.

Contact information:
School of Languages, Linguistics, Literatures and Cultures
University of Calgary
C216 Craigie Hall
Calgary, AB T2N 1N4
Canada
mgobrien@ucalgary.ca

REFERENCES

Audacity Team. (2018). *Audacity.* [Computer program]. Version 2.3.0 retrieved 11 Nov 2018 from https://www.audacityteam.org/.

Beddor, P. S., & Gottfried, T. L. (1995). Methodological issues in cross-language speech perception research with adults. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 207–232). Baltimore: York Press.

Boersma, P., & Weenink, D. (2018). *Praat: Doing phonetics by computer.* Available at www.praat.org (Last accessed November 11, 2018).

Bond, Z. S., & Small, L. H. (1983). Voicing, vowel, and stress mispronunciations in continuous speech. *Perception & Psychophysics, 34*(5), 470–474.

Caspers, J. (2010). The influence of erroneous stress position and segmental errors on intelligibility, comprehensibility and foreign accent in Dutch as a second language. *Linguistics in the Netherlands, 27*, 17–29.

Chen, H. C. (2013). Chinese learners' acquisition of English word stress and factors affecting stress assignment. *Linguistics and Education, 24*, 545–555.

Correia, S., Butler, J., Vigario, M., & Frota, S. (2015). A stress 'deafness' effect in European Portuguese. *Language and Speech, 58*(1), 48–67.

Derwing, T. M., & Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching, 42*(4), 476–490.

Domahs, U., Plag, I., & Carroll, R. (2014). Word stress assignment in German, English and Dutch: Quantity-sensitivity and extrametricality revisited. *The Journal of Comparative Germanic Linguistics, 17*(1), 59–96.

Dupoux, E., Pallier, C., Sebastian, N., & Mehler, J. (1997). A distressing 'deafness' in French? *Journal of Memory and Language, 36*, 406–421.

Dupoux, E., Sebastian–Galles, N., Navarrete, E., & Peperkamp, S. (2008). Persistent stress "deafness": The case of French learners of Spanish. *Cognition, 106*(2), 682–706.

Dupoux, E., Peperkamp, S., & Sebastian-Galles, N. (2010). Limits on bilingualism revisited: Stress 'deafness' in simultaneous French-Spanish bilinguals. *Cognition, 114*(2), 266–275.

Féry, C. 1993. *German intonational patterns.* Tübingen: Max Niemeyer Verlag.

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

Field, J. (2008). Revising segmentation hypotheses in first and second language listening. *System, 36,* 35–51.

Guion, S. G., Clark, J. J., Harada, T., & Wayland. R. P. (2003). Factors affecting stress placement for English nonwords include syllabic structure, lexical class, and stress patterns of phonologically similar words. *Language and Speech*, *46*(4), 403–427.

Grosjean, F. (1983). How long is the sentence? Prediction and prosody in the on-line processing of language. *Linguistics*, *21*, 501–529.

Grosjean, J., & Hirt, C. (1996). Using prosody to predict the end of sentences in English and French: Normal and brain damaged subjects. *Language and Cognitive Processes, 11,* 107–134.

Hahn, L. (2004). Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly*, *38*(2), 201–223.

Jarmulowicz, L., Taran, V. L., & Hay, S. E. (2008). Lexical frequency and third-graders' stress accuracy in derived English word production. *Applied Psycholinguistics*, *29*(2), 213–235.

Kleber, F., & Niebuhr, O. (2010). Semantic-context effects on lexical stress and syllable prominence. In *Speech Prosody 2010*, paper 829. Available at https://isca-speech.org/archive/sp2010/papers/sp10_829.pdf (Last accessed November 11, 2018).

Krifka, M. (2008). Basic notions of information structure. *Acta Linguistica Hungarica, 55,* 243–276.

Ladd, D. R. (1980). *The structure of intonational meaning: Evidence from English.* Bloomington: Indiana Press.

Lord, G. (2001). *The second language acquisition of Spanish stress: derivational, analogical or lexical?* (Unpublished doctoral dissertation). The Pennsylvania State University, State College, PA.

Maczuga, P., O’Brien, M. G., & Knaus, J. (2017). Producing lexical stress in second language German. *Die Unterrichtspraxis / Teaching German,50*(2), 120–135.

O’Brien, M. G., & Gut, U. (2011). Phonological and phonetic realisation of different types of focus in L2 speech. In M. Wrembel, M. Kul, & K. Dziubalska-Kolaczyk (Eds.), *Achievements and perspectives in SLA of speech: New Sounds 2010 Volume 1* (pp. 205–216). Frankfurt am Main: Peter Lang Verlag.

O’Brien, M. G., & Jackson, C. (2013, April). Realization of information structure in German speech. Paper presented at the Germanic Linguistics Annual Conference-19, University at Buffalo, Buffalo, NY.

O'Brien, M. G., & Féry, C. (2015). Dynamic localization in second language English and German. *Bilingualism: Language and Cognition*, *18*(3), 400–418.

Peperkamp, S., Vendelin, I., & Dupoux, E. (2010). Perception of predictable stress: A cross-linguistic investigation. *Journal of Phonetics, 38*(3), 422–430.

Skopeteas, S., Fiedler, I., Hellmut, S., Schwarz, A., Stoel, R., Fanselow, G., Féry, C., Krifka, M. (2006). Questionnaire on Information Structure. Potsdam: Universitätsverlag Potsdam https://www.sfb632.uni-potsdam.de/quis.html (Last accessed November 11, 2018)

Strange, W., & Shafer, V. L. (2008). Speech perception in second language learners: The re-education of selective perception. In J. G. Hansen Edwards & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 153-191). Amsterdam: John Benjamins.

Tight, D. G. (2007). Lexical subregularities and the stress preferences of L2 Spanish learners. *Hispania, 90*(3), 565–578.

Tremblay, A. (2008). Is second language lexical access prosodically constrained? Processing of word stress by French Canadian second language learners of English. *Applied Psycholinguistics, 29,* 553–584.

Tremblay, A. (2009). Phonetic variability and the variable perception of L2 word stress by French Canadian listeners. *International Journal of Bilingualism, 13*(1), 35–62.

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

van Heuven, V. J. (2008). Making sense of strange sounds: (Mutual) intelligibility of related language varieties. A review. *International Journal of Humanities and Arts Computing*, *2,* 39–62.

Yoon, T.-J., & Heschuk, J. (2011). Acoustic analysis of a Canadian French speaker's production of English stress. *Language Research*, *47*(1), 1–21.

Yu, Y-J. (2008). *A cross-language study in perception of lexical stress in English.* (Unpublished PhD dissertation). Wayne State University, Detroit, MI.

---

[i] The German carrier phrase in Domahs et al. (2014) was *Ich habe gehört, dass Peter ________ gesagt hat*. ('I heard that Peter said _______.')

[ii] An individual can open a sound file and take measurements (e.g., a vowel's formant values or duration, an utterance's peak F0) by highlighting a specific segment, word, or phrase and writing down the measurements. This is possible without creating and saving an accompanying TextGrid; however, this means that a researcher has no permanent record of precisely what was analyzed. TextGrids provide a researchers with a reliable and efficient means to re-analyze a given speech sample if necessary. *Praat*'s help menus and website provide instructions for analyzing speech samples.

[iii] The *Praat* users' group can be found at https://uk.groups.yahoo.com/neo/groups/praat-users/info.

Nagle, C. (2019). An introduction to fitting and evaluating mixed-effects models in R. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Ames, IA, September 2018 (pp. 82-105). Ames, IA: Iowa State University.

RESEARCH WORKSHOP

AN INTRODUCTION TO FITTING AND EVALUATING MIXED-EFFECTS MODELS IN R

Charles Nagle, Iowa State University

Mixed-effects modeling is a multidimensional statistical analysis capable of modeling complex relationships between predictor and outcome variables while accounting for random variance in various dimensions of the data. Although this technique is gaining popularity in applied linguistics research, learning how to model, and how to do so in R, can be intimidating. This guide provides an introduction to fitting mixed-effects models in R (Version 3.5.3) using RStudio. It includes a written introduction describing the modeling process, a video tutorial that focuses on getting started in RStudio, a sample data set, and an R script containing code to analyze the data. By the end of this introduction, researchers should have developed a basic understanding of the modeling process and should be able to (1) read data into R and inspect its structure, (2) create a series of plots to visualize trends and/or primary variables, and (3) fit and evaluate models.

INTRODUCTION

Mixed-effects models (also known as hierarchical or multilevel models) involve two fundamental components: fixed effects and random effects. Fixed effects are variables whose levels are defined, or do not change from one study to another. For example, lexical stress is a fixed effect because the levels are always reproducible across studies: stress vs. unstressed syllable, primary vs. secondary stress, etc. Gender is another classic example of a fixed effect since it is traditionally treated as a binary predictor: male vs. female. In contrast, random effects change across studies because they represent a random sample of a larger set. For instance, participants are treated as a random effect because the levels change across studies; participant 1 in study *x* is not the same person as participant 1 in study *y*. Raters and items can also be treated as random effects.

Mixed-effects models are superior to traditional analyses, such as ANOVA. Imagine that we collect data from 30 participants over four sessions, but at the last session, only 15 of our participants return. ANOVA employs listwise deletion (cases with missing data are excluded), leaving us with an analyzable sample of 15 participants; in other words, we lose all of the data for the 15 participants who completed the first three sessions. In contrast, mixed-effects models are robust in the face of missing data, which means that models are estimated using all available data points, even if some cells are missing (e.g., session 4 for 15 of 30 participants). This makes mixed-effects modeling an ideal approach for complex studies where attrition can be an issue, such as studies involving multiple tasks, sessions, or both. Mixed-effects models are also more flexible than ANOVA in terms of the assumptions that must be met. For example, the assumption of independence of observations is not required for mixed-effects models. Mixed-effects models are specifically designed to deal with dependent observations since we can treat various facets of the model as nested within one another. Finally, mixed-effects models allow for far more complex analyses, such as modeling curvilinear development and the effect of time-varying predictors, predictors whose values change over time (e.g., language use, motivation, etc.).

HOW TO USE THIS GUIDE

I began learning how to model five years ago. At the time, I did not know anything about R, so I actually learned how to use R as I was learning to model. From my perspective, learning how to model in R is remarkably similar to learning another language: as you learn how to model, you gradually restructure your knowledge, leading to a deeper and more intuitive understanding of the process, and as your familiarity increases, you become ready to learn about more complex topics. In fact, I still learn something new about modeling every time I fit models to a new data set. I share this information because I think it is important that you look at learning how to model in R as a longer-term endeavor whose payout will increase over time. This guide can serve as a starting point, but you will need to consult other resources and begin modeling your own data as soon as possible; as far as modeling is concerned, experience really is the best teacher. At the same time, I have tried to make the process as straightforward and anxiety-free as possible. In this guide, you will find step-by-step instructions on how to fit models to a sample data set using an annotated R script that I have provided. In other words, you will not need to write your own code at this stage. I have also recorded a video tutorial that will help you with preliminary steps, including setting up R and RStudio. I recommend watching the tutorial and reading this guide before modeling the accompanying data set.

Before you begin, you will need to install the latest version of R (https://www.r-project.org/) and RStudio (https://www.rstudio.com/). When you launch RStudio, it will automatically load R and ask you to create a new project in a new working directory. If you have already downloaded RStudio and created one or more projects, it may load an existing project. I prefer to create a new working directory for each project, saving all associated files (e.g., the master project files, R scripts, datasets, plots, etc.) into the folder. The written guide starts from loading the dataset and therefore assumes you have already loaded RStudio and created a new project. The video tutorial starts from opening RStudio, creating a project, and opening the script with the R code for data analysis. I will include a few illustrative screenshots of the RStudio interface in this written guide, but for information on where to click, see the video tutorial. Materials for this workshop can be accessed at https://iastate.box.com/s/bf0kerv0g17jnmqsdxgofgzldyo0ubgf.

TRANSLATING A STUDY INTO A MIXED-EFFECTS MODEL

We are going to use a data set similar to the one described in Nagle (2017). In that study, I was interested in how learners' production of L2 stop consonants changed over time. I created a set of fictitious characters to control for the phonetic context in which the stop appeared and participants' familiarity with the target items. The four fictitious characters relevant to the present analysis are *Pafo*, *Bafo*, *Pamuso*, and *Bamuso*. In the first two characters, *Pafo* and *Bafo*, the stop occurs in a stressed syllable, whereas in *Pamuso* and *Bamuso*, the stop occurs in an unstressed syllable. The outcome variable was voice onset time (VOT), an acoustic measure that represents the time that elapses between voicing onset and the release of the stop closure.

For the purpose of modeling, we will work with a data set consisting of 24 L1-English university students that I recruited from various sections of a second-semester Spanish course. Some participants had taken Spanish classes in elementary school and high school and were placed into the second-semester course, whereas others had begun learning Spanish at university. Learners participated in five sessions over their second, third, and fourth semesters of Spanish, at approximately half-semester intervals. At each session, they completed a sentence formation task and a reading task. On the sentence formation task, they saw pictures representing one of the characters, a verb, and an object or location. Using these pictures, they created simple sentences in Spanish, such as *Pafo corre en el parque* ('Pafo runs/is running in the park'). On the reading task, they saw a similar sentence printed on the computer screen and read it aloud. Ten sentences were elicited for each target character.

From this point forward, variables will appear in italics. We have the following variables in the "VOT data final.csv" data set (levels are labeled as they appear in the data set):

- *id*: categorical, 24 levels (one per participant)
- *session*: continuous, 0 to 4 (could also be treated as a factor if sessions were not evenly spaced)
- *task*: categorical, two levels (formation, reading)
- *stress*: categorical, two levels (stressed, unstressed)
- *phone*: categorical, two levels (b, p)
- *item*: categorical, four levels (this is a dummy variable that shows the character names)
- *age of learning (aol)*: continuous, the age at which the participant began learning Spanish
- *previous experience (pe)*: continuous, the number of years of Spanish participants had taken before the study
- *class*: categorical, two levels (a, b), this is a variable I have added to the data set to illustrate the principle of nesting (i.e., this variable was not part of the original study)
- *vot*: continuous, dependent variable

Before we translate this design into mixed-effects models, we should review some facts about stop consonant VOT. In English, word-initial voiceless stops such as /p/ are aspirated, or produced with a strong burst of air that delays the onset of voicing in the following segment for about 30 to 60 milliseconds depending on point of articulation (closer to 30 for bilabial /p/ and 60 for velar /k/). In contrast, voiceless stops in Spanish are unaspirated, which means that the delay between the release of the stop and the onset of voicing in the following segment is very short, ranging from 10 to 30 milliseconds. Consequently, English speakers need to minimize the burst of air that occurs on stop release to produce more Spanish-like voiceless stops. Voiced stops in English, such as /b/, are variably realized as either voiced or voiceless unaspirated. In Spanish, voiced stops are always voiced, so English speakers who produce voiced stops need to learn that only voiced variants are used in Spanish, and speakers who do not produce voiced stops need to acquire them. English speakers need to produce shorter VOT values for Spanish /p/ and negative VOT values for Spanish /b/, since negative VOT is a coding convention that indicates that voicing begins before the stop is released (i.e., that the stop is produced with voicing during closure).

Our modeling will focus on VOT in Spanish /p/, a continuous outcome whose lower and upper limits are approximately 10 and 100 milliseconds. We will focus on modeling development over time, or how *vot* changes over *session*, while examining how *task* and *stress* affect *vot*. We will also incorporate *aol* and *pe* as covariates to control for their potential relationship with *vot* (e.g., perhaps learners with an earlier *aol* produce more accurate *vot*). We could also investigate whether learners improve their VOT production more rapidly on one task, which would involve a *session* × *task* interaction.

We can model these relationships as fixed and random effects. Fixed effects essentially represent the group trend, and random effects can be conceptualized as individual variation around that trend. For instance, if we include *session* as a fixed effect, we are modeling rate of change in *vot* for a prototypical participant, pooling all of the individual data. If we include *session* as a by-subject random effect, we are modeling individual variation in rate of change. In other words, we are instructing our model to estimate a unique rate of change for each individual in the data set, or each of our 24 participants. In principle, we could include a random effect for each of our fixed effects, creating a maximal random effects structure (Barr, Levy, Scheepers, & Tily, 2013). Likewise, we could include another group of random effects, such as by-item random effects. The design of the current study includes only two target items for /p/, *Pafo* representing the stressed condition and *Pamuso* representing the unstressed condition. If we had ten target items per condition, then we could introduce by-item random effects to account for random variance associated with the particular target items we had selected.

PREPARING DATA FOR MODELING

With the conceptual basis of our model in place, we can now inspect, analyze, and plot the data in R. I recommend that you open the R script "Script for PSLLT Proceedings Article" so that you can follow each of the steps outlined below. All R code provided in this written guide will appear in Calibri. In addition to the baseline R packages, we will need the following packages: "lme4" (Version 1.1-21; Bates, Maechler, Bolker, & Walker, 2014) to fit the models, "lmerTest" (Version 3.1-0; Kuznetsova, Brockhoff, & Christensen, 2017) to produce *p* values for fixed effects, and "ggplot2" (Version 3.1.0; Wickham, 2016) to plot the data.

To install these packages, we use the `install.packages()` function. We could install each package separately using three commands, or we can install all of them simultaneously by telling R that we have multiple items, which is generally what `c()` does in R code.

```
install.packages(c("lme4", "lmerTest", "ggplot2"))
```

Having installed the packages, we now need to load or activate them using the `library()` function. In this case, we must do so individually; we cannot load all three packages simultaneously using `c()`.

```
library(lme4)
library(lmerTest)
library(ggplot2)
```

As we build our models and plots, we will create objects in R. The text that appears to the left of the arrow (which is actually the less than sign and a dash) is the name we are giving the object, and the text that appears to the right of the arrow refers to the function(s) that we are executing to create that object. First, we need to read our data into R using the `read.csv()` function.

```
data <- read.csv("VOT data.csv", row.names = NULL)
```

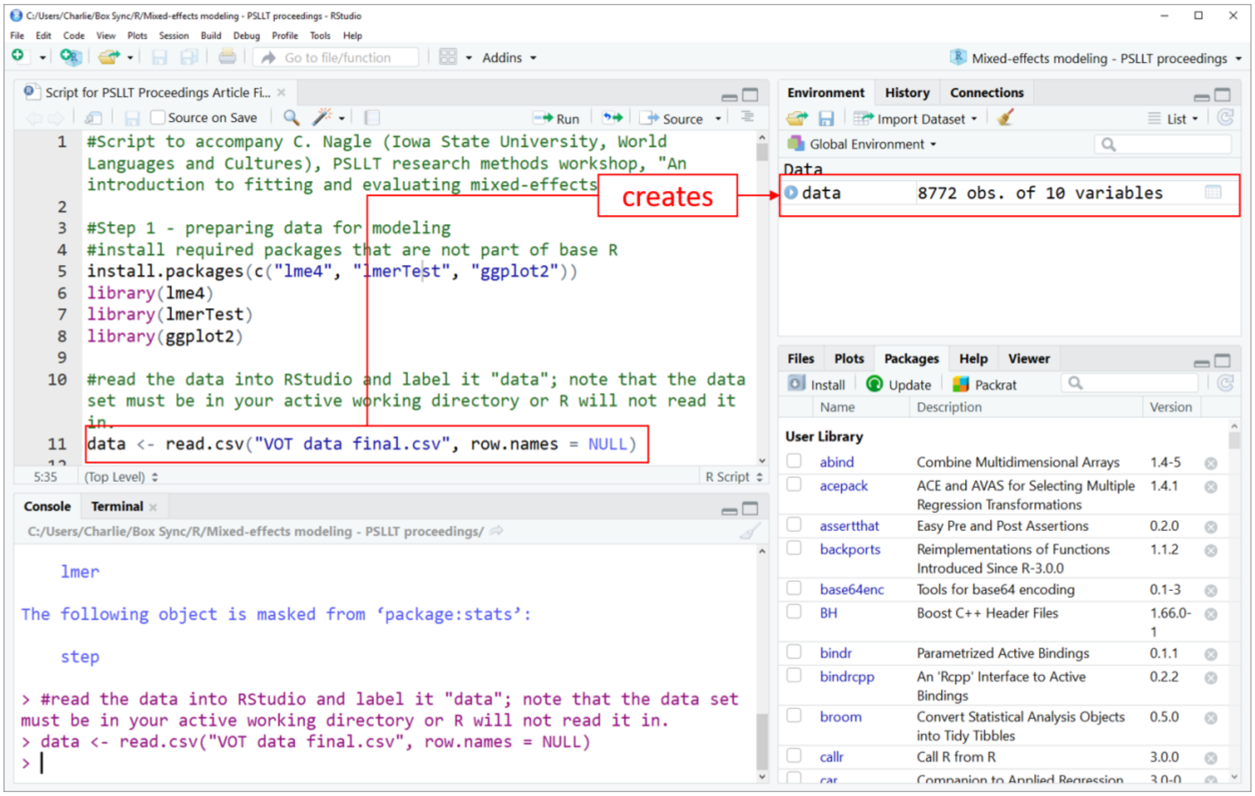


Figure 1. Screenshot of RStudio interface showing "data" dataframe on the Environment tab. The font, size, and background of the interface will depend on your settings in tools > global options > appearance.

We now have a dataframe named "data" (Figure 1) that we can inspect using the str() function. We should inspect every dataframe to make sure that R has interpreted our data structure properly. In my research, I typically use numbers to refer to participants (e.g., 1, 2, 3, etc.). R interprets numbers as integers or continuous variables. To avoid this, you could label participants with letters or a combination of a letter and number (i1, i2, i3, etc.). However, if you like to use numbers like I do, then we can use the as.factor() function to tell R that *id* is a categorical variable. This function is slightly more complicated since we need to use $ to tell R to look inside the dataframe for the *id* variable and interpret it as a factor.

```
str(data)
data$id <- as.factor(data$id)
```

Now if we reinspect the data using the str() function, we see that R is interpreting *id* as a factor. All of our other variables have been interpreted correctly. We are going to focus on the /p/ data for this analysis, so we need to create a new data set consisting of only the /p/ data using the subset() function.

```
data.p <- subset(data, phone == "p")
```

From this point forward, we will work with the “data.p” dataframe. The last step before we begin analyzing is to check for normality using the qqnorm() function. If our continuous variable is normally distributed, the points will fall more or less on a straight diagonal line. While slight deviations are acceptable, major departures indicate that the distribution is not normal.

```
qqnorm(data.p$vot)
```

Immediately from the plot we can see that some participants produced negative VOT (or voiced variants) for /p/. For these target items, participants probably made a mistake, reading /p/ as /b/, so it makes sense to eliminate these few outliers from the data set and retest for normality. We can use subset() to remake the data set including only items for which VOT > 0. In this case, we are modifying (overwriting) our dataframe instead of creating a new one, so we include “data.p” to the left of the arrow.

```
data.p <- subset(data.p, vot > 0)
```

If we look at the environment tab, typically displayed in the upper right corner of RStudio, we can see that the number of observations decreased from 4392 to 4375 when we executed the subset function, which tells us that we have eliminated 17 observations whose VOT ≤ 0. Now we check for normality again, and see that the data is beginning to look more normal since we have eliminated outliers on the lower end. However, we can still see that there is a relatively substantial curve in the line, so we will transform the data to enhance normality. Before we do, we should look at histogram of *vot*, which can help us determine what type of transformation to apply. We can generate a plot object (Figure 2) using the following code:

```
histogram.p <- ggplot(data.p, aes(x = vot)) +
    geom_histogram()
```

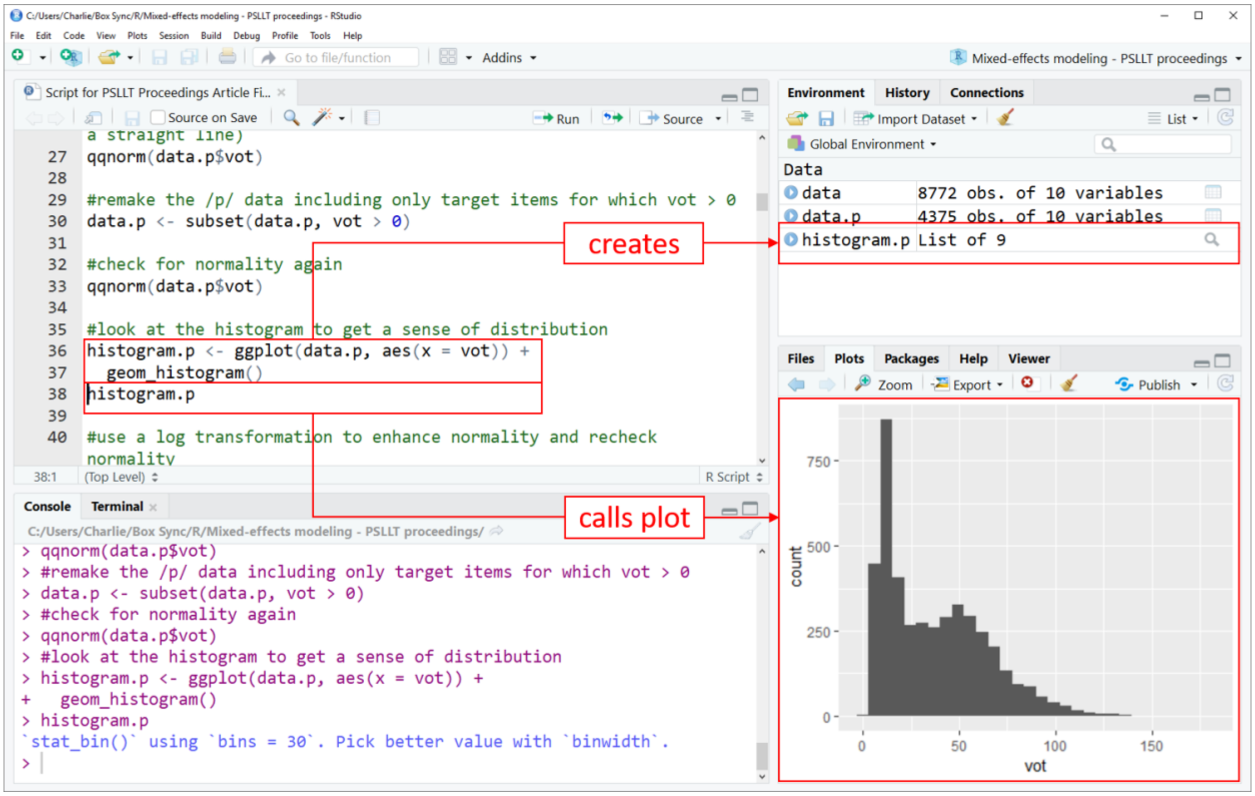


Figure 2. Screenshot of code to generate “histogram.p” plot and code to call the plot in the “Plots” tab.

The histogram seems to suggest that the data is somewhat log-normal, making a log transformation appropriate. Consequently, we will add a new variable, *log_vot*, to our dataframe using the log() function.

```
data.p$log_vot <- log(data.p$vot)
```

Now when we check normality of *log_vot*, qqnorm(data.p$log_vot), we see that we have a reasonably straight line. We will build models using *log_vot* as our dependent variable.

VISUALIZING THE DATA

When we model, we are trying to represent the data as accurately as possible, assessing relationships between our predictors and outcome. Plotting provides insight into the data and helps us fit the appropriate model, particularly when we are dealing with complex data sets. For longitudinal data in particular, we are trying to visualize the shape of the developmental curve so that we can specify the time predictor appropriately. For instance, if we see that development slows down over time, then we could include linear and quadratic time variables to estimate linear and quadratic rates of change (i.e., rate of change and rate of deceleration).

I always generate at least two primary plots, one that summarizes the group trajectory against individual trajectories (plots that include individual trajectories are sometimes referred to as spaghetti plots) and another that plots trajectories for individual participants over time. We will use the “ggplot2” package to make our plots. Code for this package can be complicated depending on what we want to graph and how we want it displayed. I recommend starting each piece of code on a new line for the sake of readability. Make sure to include a plus sign (+) at the end of each line, except the last, so that R continues reading the next piece of code. If you take this approach, R will automatically indent lines below the first to indicate that indented lines pertain to a larger block of code, as in the following example.

```
plot.group.p  <- ggplot(data.p, aes(session, vot)) +
    stat_smooth(method = "loess", se = F, size = 2) +
    stat_summary(aes(group = id), fun.y = mean, geom = "line", alpha = 0.3) +
    xlab("Session") +
    ylab("Voice Onset Time")
```

This creates the basic plot. However, I like to remove the gridlines and use a black and white theme, which we can accomplish by adding the following pieces of code. The two sets of code will generate the plot displayed in Figure 3.

```
    theme_bw() +
    theme(strip.background = element_blank()) +
    theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank())
```

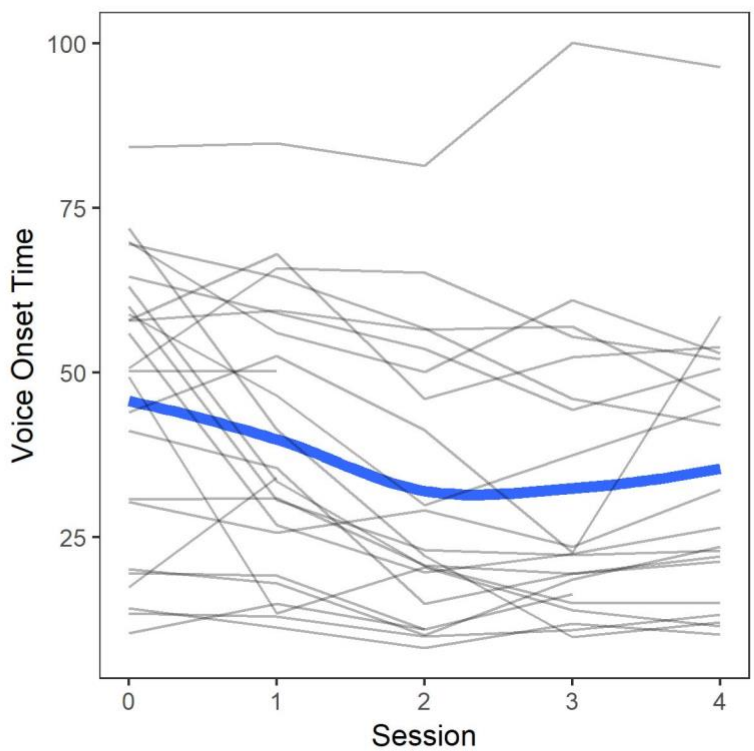


Figure 3. Group trend in VOT for Spanish /p/ versus individual trajectories.

If we want to save our plot, we use the ggsave() function, specifying the output file, the plot object in R, and the dimensions of the plot in inches and its dpi (300 is typically required by most journals).

```
ggsave("group plot p data.jpeg", plot.group.p, width = 4, height = 4, dpi = 300)
```

We can rework the code for the group plot to make individual boxes for each participant by telling R to facet (or array the data) by *id* using the facet_wrap() function. We can also optionally include the number of columns and rows if we want a particular configuration. In this case, we will specify six columns (generating four rows) since we have 24 participants.

```
plot.individual.p  <- ggplot(data.p, aes(session, vot)) +
    stat_summary(fun.y = mean, geom = "line") +
    xlab("Session") +
    ylab("Voice Onset Time") +
    theme_bw() +
    theme(strip.background = element_blank()) +
    theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank()) +
    facet_wrap(~id, ncol = 6)
```

We then save this plot using the ggsave() function, adjusting the dimensions to fit the plot. In general, getting the dimensions of the plots right requires some trial and error, so I think the simplest approach is to save the file, copy it into a Word document, and adjust the dimensions as needed until it displays correctly. For example, I specified dimensions of 6.5" width and 4" height to create Figure 4.

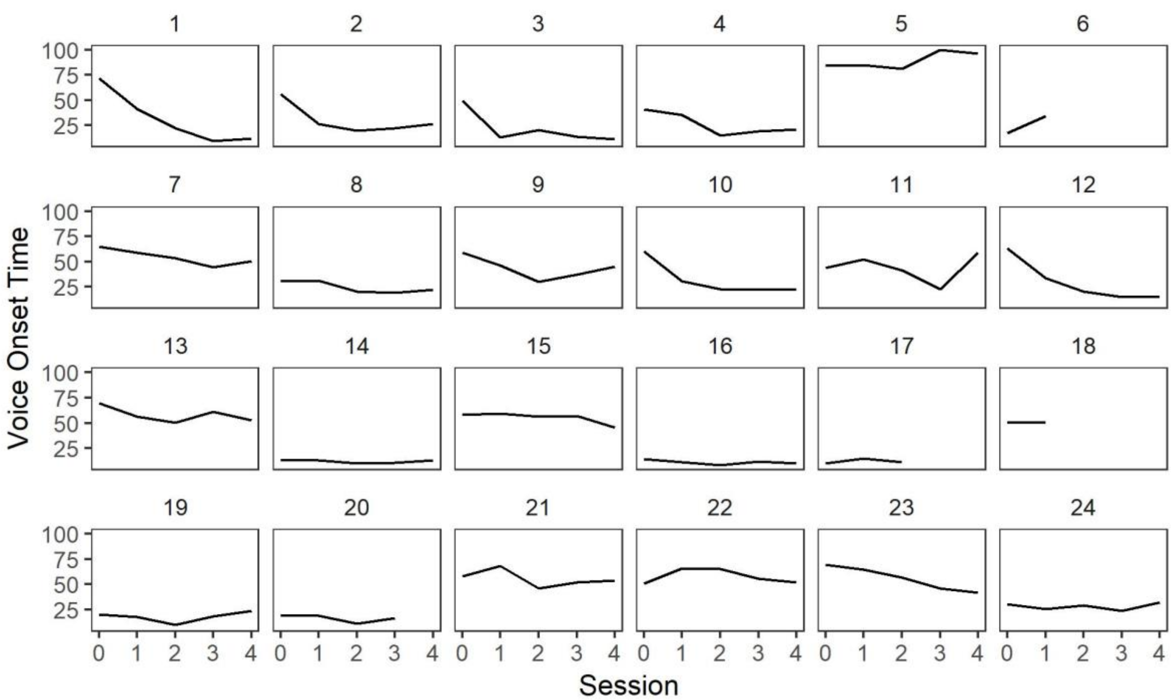


Figure 4. Individual trajectories in VOT for Spanish /p/.

These two plots (Figures 3 and 4) show us that most participants improved their VOT production since we see a downward trend toward shorter, more Spanish-like VOT in /p/. We also see that for many individuals, VOT improved more over the first half of the study, suggesting that we should try modeling linear and quadratic rates of change. In both plots, we can see that some learners did not participate in all five sessions (e.g., 6, 18). As I mentioned in the introduction, mixed-effects models can handle missing data, so these cases are not problematic.

We could also generate a plot to illustrate the effect of stress on VOT production. In the code below, we map line type to stress (we could also map color, but since many journals print in greyscale, I try to avoid using color to differentiate conditions or groups) within the aes() function and include an additional line of code to move the legend from its default location (vertical display to the right of the plot) to the bottom of the plot to avoid compressing the x-axis.

```
plot.individual.p  <- ggplot(data.p, aes(session, vot, linetype = stress)) +
    stat_summary(fun.y = mean, geom = "line") +
    xlab("Session") +
    ylab("Voice Onset Time") +
    theme_bw() +
    theme(strip.background = element_blank()) +
    theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank()) +
    facet_wrap(~id, ncol = 6) +
    theme(legend.position = "bottom")
```

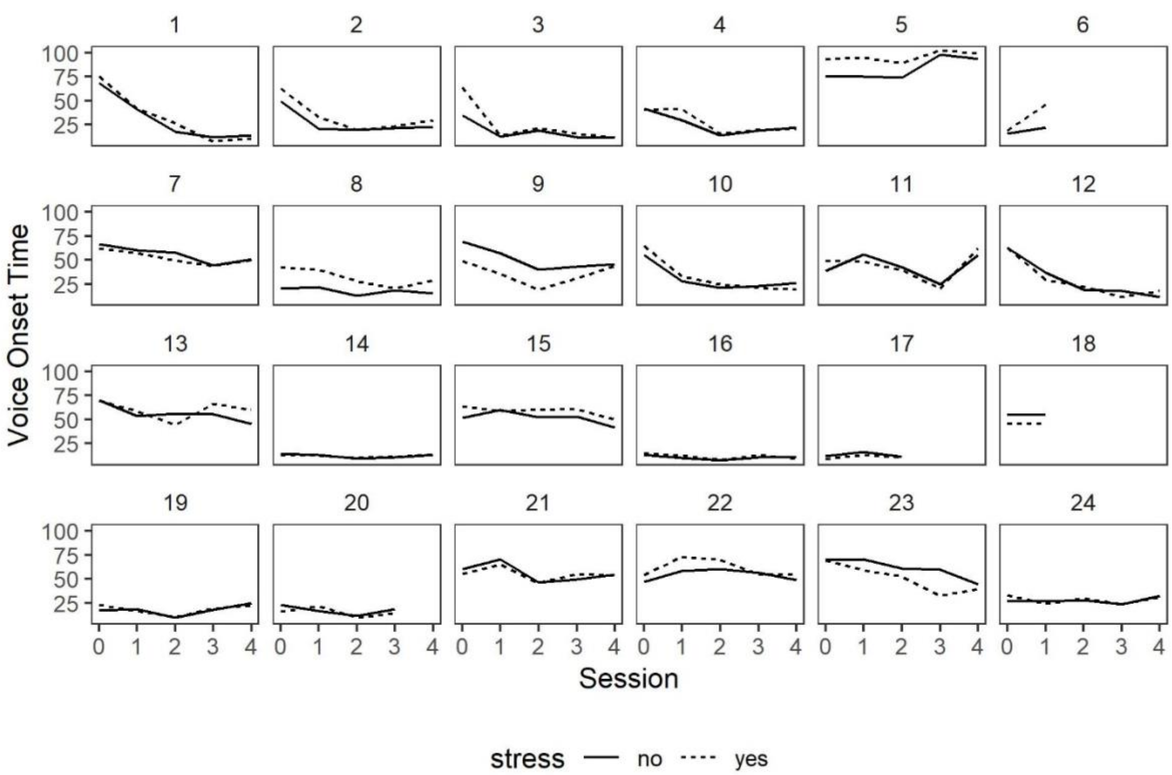


Figure 5. Individual trajectories in VOT for Spanish /p/ in stressed and unstressed environments.

FITTING MIXED-EFFECTS MODELS TO THE DATA

Before we can begin modeling, we should create a variable that represents the quadratic trend for time by squaring *session*. We want to create this variable because in Figures 3 and 4 we observed curvature in the group and individual trajectories. In other words, VOT production over time did not follow a straight line (cf. participants 1, 2, 10, 12, etc. in Figure 4 above). We can approximate this curvature using a quadratic function for *session*, or *session_sq*.

```
data.p$session_sq  <- data.p$session^2
```

We should also center *aol* and *pe*, our continuous predictors. Centering is essentially a form of standardizing variables, making them easier to interpret without altering the model (coefficients will change but significant effects will not). There are various ways to center, but grand-mean centering makes the most sense for our predictors. In this form of centering, we compute the mean and subtract it from each participant's score on the relevant variable. In our centered variables, a negative score indicates that the participant was below the mean, a positive score that the participant was above the mean, and a score of zero refers to the mean. For example, *aol* refers to the age at which participants began learning Spanish. When we fit a model, the intercept is computed by setting all predictors to zero, but a score of zero is not possible in our data (i.e., a score of zero would in theory represent a native speaker of Spanish). Centering resolves this issue by setting zero to refer to the sample mean. A score of zero is possible for *pe* because some participants had not taken Spanish before enrolling in university language coursework. However, it is still advantageous to center *pe* to represent the average amount of previous experience that participants had, since ultimately we are trying to model a prototypical participant's trajectory. We can create the centered predictors and add them to our data set in R.

```
data.p$aol_c  <- data.p$aol – mean(data.p$aol)
data.p$pe_c  <- data.p$pe – mean(data.p$pe)
```

We are now in a position to begin modeling. Scholars have advocated for a variety of approaches to modeling, but the most common is forward-testing random effects and backward-testing fixed effects. This means that we will add random effects one by one and compare models to one another, and we will add fixed effects as a group and compare models by progressively dropping the least significant effects. Even though backward-testing fixed effects is generally advisable, this does not mean that we should include every possible fixed effect. Rather, our fixed effects should be guided by our theoretical framework and the design of our study. In certain scenarios, we may decide to retain a fixed effect even if is not significant. For instance, if we are interested in higher order interactions among predictors, we would not eliminate their baseline components.

In practice, backward-testing fixed effects can be challenging when you first transition to mixed-effects models. In my view, a data-driven, bottom-up approach (Cunnings & Finlayson, 2015) is perfectly acceptable so long as you report your modeling process as transparently as possible, including the model containing your "final" set of predictors. The data-driven approach can be particularly advantageous when dealing with longitudinal data since decisions need to be made about the shape of development over time. In some of my previous work (Nagle, 2017a, 2017b), I opted to forward-test models, reporting all of the models I built in a table or appendix to illustrate the process (for examples of how to format such a table, see Murakami, 2016; Singer & Willett, 2003).

For the purpose of this illustration, we will take a hybrid approach. First, we will build the unconditional growth model, which is a model that describes how VOT, our dependent variable, changes over time. To this model, we will add task and stress, our primary predictor variables, using backward-testing to evaluate the fixed terms and forward-testing to evaluate the random terms. To build our models we will use the lmer() function, which takes the following general form. Note that I place fixed and random effects on separate lines so that the code is easier to read.

```
name of model  <- lmer(dependent variable ~ fixed1 + fixed2 +
    (random1 + random2 | random grouping term 1) +
    (random1 + random2 | random grouping term 2), data = name of dataframe)
```

First, we build the null or random intercepts model:

```
null.p  <- lmer(log_vot ~ 1 + (1 | id), data = data.p, REML = F)
```

In the code above, we are creating a model, "null.p," in which *log_vot* is the dependent variable. We have only one fixed effect, the intercept, represented by the *1* after the tilde (~), and we have included by-subject random intercepts using the code (1 | id). In this piece of code, the random effects appear to the left of the vertical bar, and the grouping term over which they are computed to the right. We have also specified that our data set is "data.p," and we have told the model to use maximum likelihood estimation rather than restricted maximum likelihood estimation (REML) so that we can compare models with different fixed-effects structures to one another. If we were interested in comparing models with the same fixed effects but different random effects, then we could fit and compare REML models. In lmer(), REML is the default, so we turn it off using the code REML = F.

Next, we build unconditional linear and quadratic growth models using *session* and *session_sq*. Unconditional growth simply means that we have not yet included any predictors that would affect the intercept or rate of change over time (i.e., we have not yet placed any conditions on the intercept or rates of change). R will always include intercepts unless we suppress them, so we do not need to carry *1* forward in our model specification.

```
linear.p  <- lmer(log_vot ~ session +
    (session | id), data = data.p, REML = F)

quadratic.p  <- lmer(log_vot ~ session +
    session_sq +
    (session + session_sq | id), data = data.p, REML = F)
```

The quadratic model with by-subject random slopes for session_sq fails to converge, which is not uncommon for models involving large data sets and/or complex random effects structures. Failure to converge does not always indicate a problem with model specificiation. There are a number of ways to facilitate convergence, such as using a different optimization function, removing covariances among random effects, and simplifying the random effects. For the purpose of this guide, we will instruct R to use the BOBYQA optimizer instead of the default nloptwrap implemented in lme4 version 1.1-20 (for more information on convergence, see the FAQ section at the end of this guide). We can do this using the following code. For the sake of readability, I will move the data and fit specifications to a separate line.

```
quadratic.p  <- lmer(log_vot ~ session +
    session_sq +
    (session + session_sq | id),
    data = data.p, REML = F, lmerControl(optimizer = "bobyqa "))
```

Now we compare the three models (null, linear growth, and quadratic growth, each with the accompanying by-subject random effects) using the anova() function. This function performs a chi-square test on the change in the deviance statistic for nested models, or models that can be derived from one another by setting one or more parameters to zero.

```
anova(null.p, linear.p, quadratic.p)
```

The output in R (Figure 6) shows that each model is an improvement over its predecessor, which is not surprising since our plotting already revealed a quadratic trend in both the group and individual data. We can also see that the change in degrees of freedom (the Df column) is 3 for the comparison between the null.p and linear.p models, and 4 for the comparison between the linear.p and quadratic.p models. This may seem odd since we only added two terms, one representing the fixed effect and one representing the random effect, to each model. When we added those terms, R also included covariances among the random effects: between *session* and the intercept for the linear.p model; between *session_sq* and the intercept, and *session_sq* and *session* for the quadratic.p model.

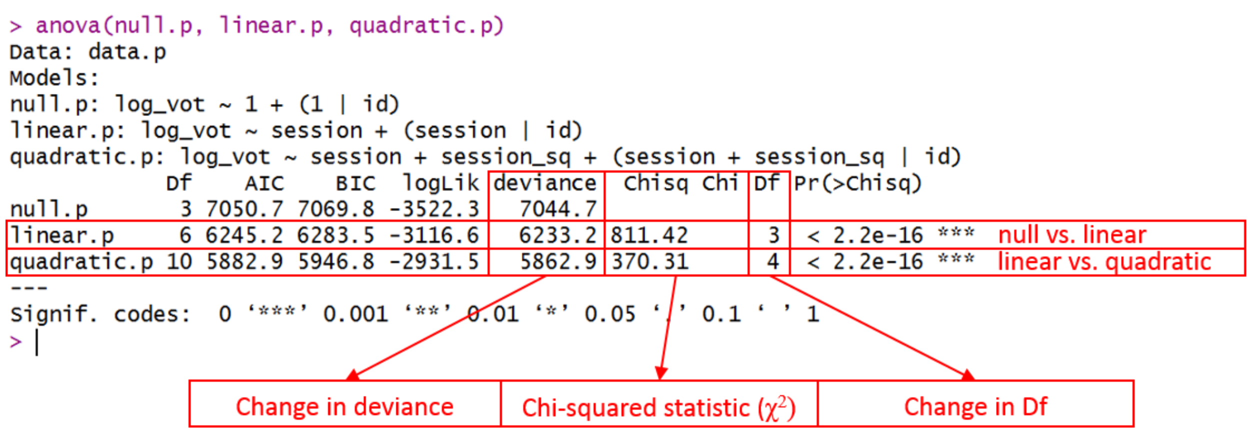


Figure 6. Screenshot of output for model comparisons using the ANOVA() function. Rows represent model comparisons (i.e., linear.p reports the null.p vs. linear.p comparison, and quadratic.p reports the linear.p vs. quadratic.p comparison).

We can request a summary (Figure 7) of our unconditional quadratic growth model using the summary() function.

```
summary(quadratic.p)
```

The first part of the summary shows the formula for the model we fit, followed by fit statistics (e.g., AIC, BIC, and deviance), and residuals. The random effects summarize the variance in intercepts, linear slopes (*session*), and quadratic slopes (*session_sq*), as well as the residual within-subjects variance in the model. The fixed effects, listed below the random effects portion of the model, demonstrate that there is a negative trajectory over the course of the study (i.e., the coefficient for *session* is negative), which in this case indicates improvement. The positive coefficient for the *session_sq*, when interpreted with respect to the negative coefficient for *session*, indicates that development decelerated over time. These findings align with the initial plots we generated. We can compare the magnitude of the coefficients and their directionality, but we must remember that we are fitting models to *log_vot*, so the coefficients do not refer to the intercept or rate of change on the original VOT scale.

```
> summary(quadratic.p)
Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's method
lmerModLmerTest]
Formula: log_vot ~ session + session_sq + (session + session_sq | id)
   Data: data.p
Control: lmerControl(optimizer = "bobyqa")

     AIC      BIC   logLik deviance df.resid
  5882.9   5946.8  -2931.5   5862.9     4365

Scaled residuals:
    Min      1Q  Median      3Q     Max
-4.9771 -0.5491  0.0001  0.5326  4.4791

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 id       (Intercept) 0.397470 0.63045
          session     0.140028 0.37420  -0.29
          session_sq  0.004208 0.06487   0.14 -0.95
 Residual             0.210629 0.45894
Number of obs: 4375, groups:  id, 24

Fixed effects:
            Estimate Std. Error       df t value Pr(>|t|)
(Intercept)  3.61756    0.12949 23.98151  27.937  < 2e-16 ***
session     -0.31554    0.07851 23.24767  -4.019 0.000528 ***
session_sq   0.05743    0.01404 21.27936   4.090 0.000512 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
           (Intr) sessin
session    -0.297
session_sq  0.149 -0.942
```

Summary of model fit

Fit statistics

Summary of random effects (by-subject, or *id*)

Summary of fixed effects (*p* values estimated using lmerTest package)

Figure 7. Screenshot of summary of quadratic.p.

We are now ready to integrate our remaining fixed-effect predictors (*stress*, *task*, *aol_c*, and *pe_c*) as a block and backwards test them. In the following code, I have included our time predictors, *session* and *session_sq*, on the same line so that you can more easily see the four new fixed effects we have added.

```
fm1.p <- lmer(log_vot ~ session + session_sq +
    task + stress + aol_c + pe_c +
    (session + session_sq | id),
    data = data.p, REML = F, lmerControl(optimizer = "bobyqa "))
```

Again, we ask for a summary of the model using the summary() function. In the list of fixed effects, *task* is labeled "taskreading" and *stress* is labeled "stressyes." R interprets categorical predictors alphabetically. Thus, the sentence formation task, "formation," and the unstressed context, "no," have been set as the baseline conditions against which the reading task and stressed context are compared (for information on contrast coding categorical predictors, see Linck & Cunnings, 2015). In other words, the intercept refers to VOT production on the formation task when the stop occurs in an unstressed syllable. If we have installed the "lmerTest" package, then summary() will return *p* value estimates for each predictor.

We can see from our summary that *task* is significant; the negative coefficient for "taskreading" indicates that on average, participants produced lower VOT values on the reading task than on sentence formation, and the large *t* value suggests that this effect was relatively robust. In contrast, *stress* was not significantly related to VOT production, and neither were our grand-mean centered covariates, *aol_c* and *pe_c*. We now have two options. Following the principle of backward-testing fixed effects, we could rank these predictors in terms of their *t* value and begin dropping them from the model in the following order: *pe_c*, *stress*, and *aol_c*. In this case, we would need to report the order in which we dropped the nonsignificant fixed effects and the corresponding anova() model comparisons at each stage. The second option would be to retain all effects, reporting our final model so that readers can more easily compare and contrast estimates, standard errors, and *t* values for all of the fixed effects included in the study. In my view, it is important to create a parismonious model that does not include a large number of nonsignificant fixed effects, but it is not always advisable to strive for the minimally adequate model, or a model that contains only those fixed effects that enhance fit (i.e., only significant fixed effects).

For our data set, we will refit the model without stress and compare the simpler model to its more complex predecessor. However, we will keep *aol_c* and *pe_c* since we have included them as control covariates.

```
fm2.p <- lmer(log_vot ~ session + session_sq +
    task + aol_c + pe_c +
    (session + session_sq | id),
    data = data.p, REML = F, lmerControl(optimizer = "bobyqa "))
anova(fm1.p, fm2.p)
```

The chi-square statistic ($\chi^2(1) = .35$, $p = .56$) indicates that including *stress* does not significantly change model fit, so we can drop it from the model. However, in discussing the models, we would still report and interpret stress since lack of significance is an important finding that should not be ignored.

Now we can focus on our random effects. Some scholars have advocated for a maximal approach, which means including a random effect for every fixed effect (Barr et al., 2013; Cunnings &

Finlayson, 2015). In my view, fitting an appropriate set of random effects can be challenging, so I would not recommend taking a maximal approach. For the sake of our model, we will add a random effect for *task* to see if it improves the model, keeping in mind that as we add more random effects, the model may take longer to converge (i.e., random effects and their covariances can be computationally intensive). When we model *task* as a by-subject random effect, we are allowing R to estimate unique coefficients for *task* for each participant. Put another way, for some participants, *task* may have had a strong impact on VOT production (i.e., large differences in VOT by task), whereas for others its effect may have been comparatively weak (i.e., small differences in VOT by task). We test this possibility by including *task* as a random effect.

```
fm3.p  <- lmer(log_vot ~ session + session_sq +
     task + aol_c + pe_c +
     (session + session_sq + task | id),
     data = data.p, REML = F, lmerControl(optimizer = "bobyqa "))
anova(fm2.p, fm3.p)
```

Including *task* as a by-subject random effect has significantly improved fit ($\chi^2(4) = 93.05$, $p < .001$), so we will keep it in the model. We can now consider this our "final" model, or the model that is the best representation of our data given our research aims and predictors. Now that we have a final model, we should calculate 95% confidence intervals for fixed effects using the `confint()` function. Profiling confidence intervals can take a very long time depending on the complexity of the model and the computer's processor. Using an older desktop with eight gigabytes of RAM, I waited nearly 15 minutes for R to produce confidence intervals before stopping `confint()`.

```
confint(fm3.p)
```

We can speed up the process by approximating the confidence intervals using the Wald method, which is far quicker (less than one second on the same machine). If we approximate the intervals, then we should report this in the manuscript.

```
confint(fm3.p, method = "Wald")
```

Table 1 reports the final model following Linck and Cunning's (2015) format, including a note to indicate that the confidence intervals are approximate. I have assigned the fixed effects (i.e., parameters) a more informative label instead of using the variable names as they appear in our data set.

Table 1

Summary of components in mixed-effects model of VOT development in L2 /p/

						Random effects
	Fixed effects					By Subject
Parameter	Estimate	*SE*	95% CI	*t*	*p*	*SD*
Intercept	3.70	.13	[3.45, 3.95]	28.91	< .001	.62
Linear slope	–.31	.08	[–.47, –.16]	–4.02	.001	.37
Quadratic slope	.06	.01	[.03, .08]	4.11	.001	.06
Task: Reading	–.17	.04	[–.24, –.10]	–4.64	< .001	.17
Age of learning	.05	.08	[–.10, .20]	.63	.53	
Previous experience	.05	.10	[–.15, .24]	.46	.65	

Note. 95% CI were approximated using the Wald method.

Finally, there are a number of ways to evaluate how well our model fits the data, such as plotting fitted against residual values. In my experience, the latter typically works well for linear relationships but can be misleading when modeling polynomial change, such as the quadratic term (*session_sq*) we included in our models. One simple alternative is to generate a set of predicted values based on our model and then compare predicted growth to observed growth over time. Including this type of plot as an appendix or supplementary file can be helpful. We can generate predicted values and add them to our data set using the predict() function.

```
data.p$predicted <- predit(fm3.p)
```

We currently have two dependent variables, *log_vot* and *predicted*, that are stored in two separate columns. If we want to plot the predicted and observed data on the same plot, we need to transform the data into a new longitudinal data set by merging the two dependent variables into a single column and creating a new identifier variable. There are many packages and approaches we could take, but I prefer the "tidyverse" package (Version 1.2.1). In the R code below, I use the gather() function. We include the data set (data.p), the name of the new identifier column (*type*, for type of data: observed vs. predicted), the name of the new outcome variable (*log_vot2*), and the variables to be merged (*log_vot* and *predicted*). We also use c() to tell R that we are dealing with multiple variables.

```
install.packages("tidyverse")
library(tidyverse)
data.predicted <- gather(data.p, type, log_vot2, c(log_vot, predicted), factor_key = TRUE)
```

We can now plot the data as before, using our code to generate individual plots while mapping *type* to line type:

```
plot.predicted.p <- ggplot(data.predicted, aes(session, log_vot2, linetype = type)) +
    stat_summary(fun.y = mean, geom = "line") +
    theme_bw() +
    theme(strip.background = element_blank()) +
    theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank()) +
    xlab("Session") +
    ylab("Log VOT") +
    facet_wrap(~id, ncol = 6) +
    theme(legend.position="bottom")
```

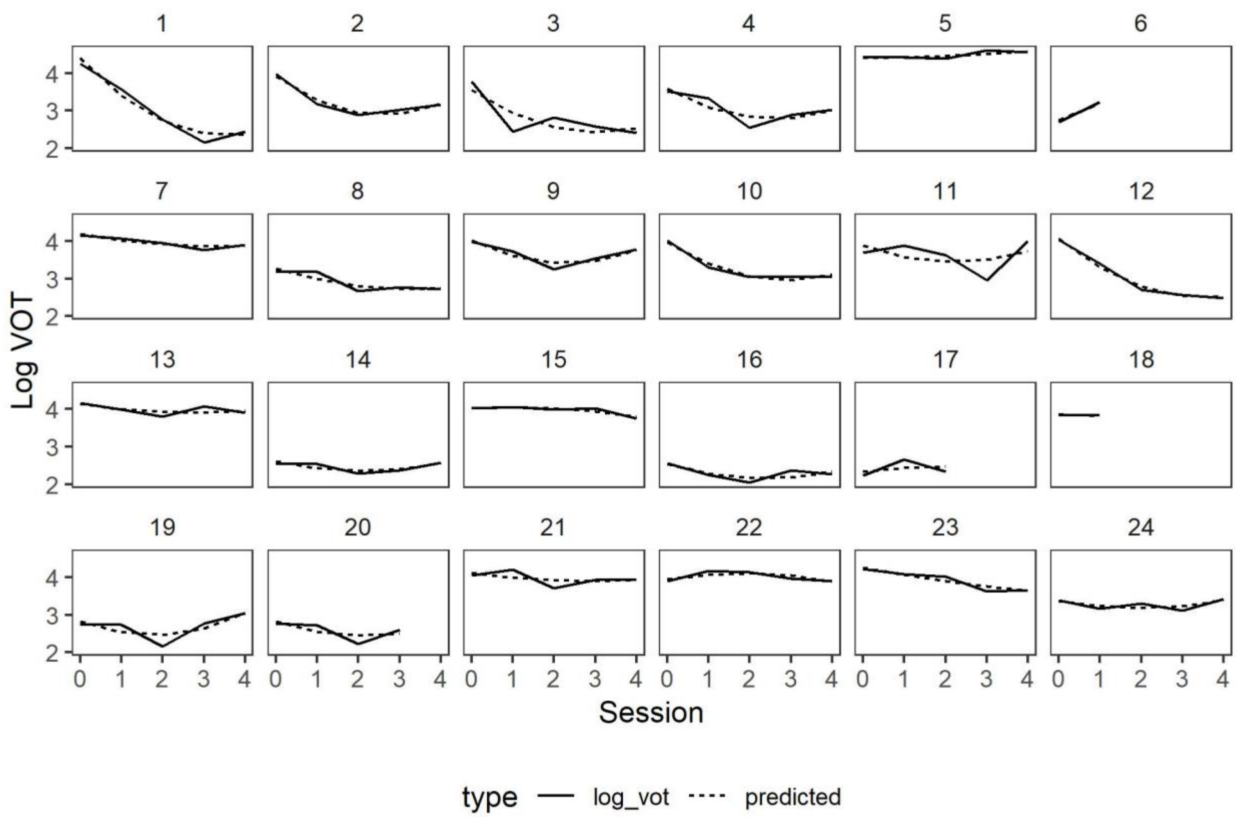


Figure 8. Observed (log_vot) vs. model-predicted individual trajectories.

Comparing the dotted lines in Figure 8, which represent the model-estimated values, we can see that each individual plot displays different rates of linear and quadratic change. This serves as a visual reminder that we instructed R to estimate unique rates of change for each participant in the data set by including *session* and *session_sq* in the random-effects structure of our model. We can also see that the model represents the data reasonably well.

One last plot that we might be interested in generating is the model-estimated group trajectory. This plot is similar to Figure 3, but we will graph the model trajectory as a dashed line. In general, I use solid lines for observed data and dashed lines for model-estimated data as appropriate.

```
plot.model.p <- ggplot(data.p, aes(session, log_vot)) +
    stat_summary(aes(y = fitted(fm3.p)), fun.y = mean, geom = "line", linetype = "dashed", size
    = 2) +
    stat_summary(aes(session, log_vot, group = id), fun.y = mean, geom = "line", alpha = 0.2) +
    theme_bw() +
    theme(strip.background = element_blank()) +
    theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank()) +
    xlab("Session") +
    ylab("Log VOT")
```

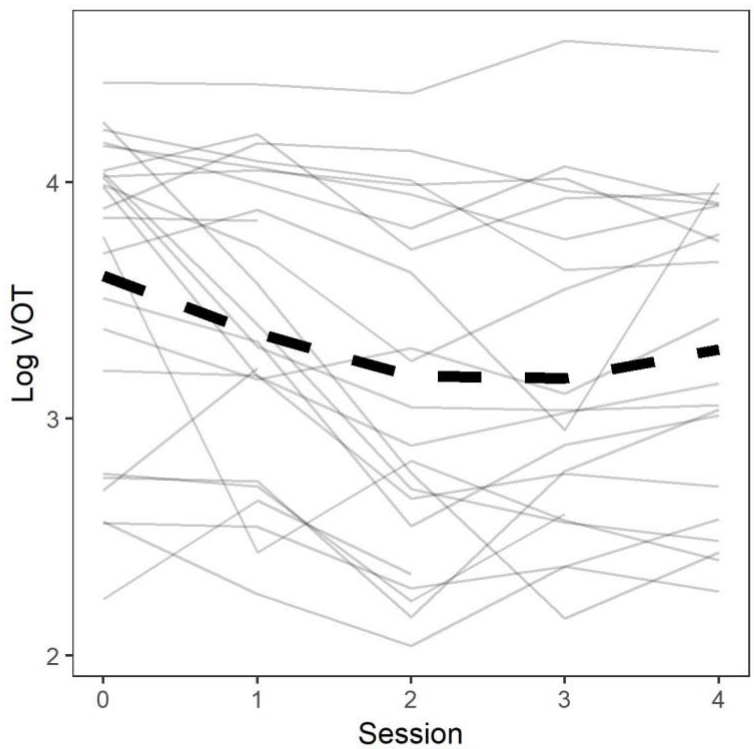


Figure 8. Model-estimated group trajectory vs. observed individual trajectories.

MODELING NESTED DATA

Up until now, we have assumed that we have drawn a random sample of students from a variety of sections of the same course, and that these students over time have had different instructors. Now we will consider another case. Let's assume that we recruited students from two different sections of Spanish, labeled a and b in the data set, and followed them over a single semester of coursework. In this scenario, we can say that the students are nested in classes. If we had a multisite design, then students would be nested in classes and classes nested in schools (the latter would be a three-level model). Modeling nesting is important because each class (or school) may display a unique growth rate, and we would expect growth rates for students in the same class to be more similar to one another than to growth rates for students in different classes (e.g., higher correlation within classes). The R code for creating nesting is a forward slash with the higher-order group first and the lower-order group second, such as class/id or school/class/id. This syntax gets included as a grouping factor for the random effects (i.e., on the right side of the vertical bar).

```
fm.p.nested  <- lmer(log_vot ~ session + session_sq +
    task + aol_c + pe_c +
    (session + session_sq + task| class/id),
    data = data.p, REML = F, lmerControl(optimizer = "bobyqa "))
```

If we generate a summary of the nested model using the summary() function, we can see that the random effects structure now includes both "id:class" and "class." We now have estimates for the variance among classes and the variance among students, taking into consideration the fact that students were nested in classes. However, when we fit the model, R returned a singular fit warning, which means that the model was over-specified. Singular fit is not surprising since in our data set we have very few observations for class ($n = 2$). Thus, estimating unique linear and quadratic slopes for each class would not be advisable. If we were interested in variance among classes, we would probably want to achieve a sample size of at least 10–20 classes, with 10–20 students in each class.

FREQUENTLY ASKED QUESTIONS

1. What if one of the variance components in the random effects structure of my model is very small?

   Interpret the variance components with respect to their corresponding fixed effect. A variance component that is very small relative to its fixed effect indicates that there is virtually no between-subjects variance in that parameter. In that case, even if the inclusion of the random effect significantly improves model fit, it may be advantageous to select a simpler model without that term. In our final model reported in Table 1, the between-subjects variance in quadratic slopes (*session_sq*) is .004. This is small relative to some of the other variance components, but proportionate relative to the fixed effect estimate (.06). For more information on model selection, see Murakami (2016).

2. What if my model fails to converge?

   More complex models may fail to converge. There are a number of solutions you can attempt. First, you can change the default optimizer, which is what we did using lmerControl(optimizer = "bobyqa"). You could also try fitting the model using all optimizers by fitting the model using the default optimizer and then running the allFit() function on your model object.

   Simplifying the random effects can also facilitate convergence. You can eliminate covariances among the random effects by including a double vertical bar: (1 + session + session_sq || id). If the model stills fails to converge, then simplifying the random effects could help (i.e., eliminating higher-order random effects such as interaction terms). For more information, see Barr et al. (2013) and Linck and Cunnings (2015).

3. What if my outcome variable is categorical?

   Linear models are appropriate for continuous outcome variables. Generalized linear models (glimmer) are appropriate for categorical outcomes. Thus, glimmer would be appropriate for perception data that are coded as correct/incorrect, for production data that are coded as intelligible/unintelligible, etc. You can fit a generalized linear model using the glmer() function. The specification for glmer() is essentially the same as lmer(), but the interpretation of glimmer is not as straightforward. For more information on glmer(), see Baayen (2008) and Linck and Cunnings (2015).

4. What about ratings data?

   When comprehensibility, accentedness, and other ratings are carried out on a 1000-point sliding scale, they can be considered continuous. In that case, the same procedure outlined above can be followed, and rater would be included as a grouping for random effects. In other words, two sets of random effects would be expected: one grouping for speakers and another for raters. For example, had this study been a ratings study, we might expect the following random effects structure.

```
(1 + session + session_sq | speaker) +
(1 + session + session_sq | rater)
```

   When ratings are carried out on a shorter scale, such as a 9-point scale, then the data is ordinal. The best approach for modeling this type of data is to pool data over raters, in which case rater would not be included as a random effect. This will linearize the scale and make it suitable for modeling with lmer().

5. What if my data is not longitudinal?

   Most of my research is longitudinal, which is why I have concentrated on modeling longitudinal data. You can model cross-sectional data following the exact same procedure, but you will not need to introduce time predictors, such as *session* and *session_sq*, into the model.

6. What if I do not know or remember the R code for a particular function or analysis?

   First, if you are confused about a particular function, you can type a ? before the function and R will give you a description of what it does, such as ?lmer. The R community is also very large. In general, you can find what you are looking for by consulting the R cookbook (http://www.cookbook-r.com/) or searching R forums. For online searching, start by including the package and/or function you are using and a short description (e.g., fitting piece-wise growth models using lmer, modifying the x-axis in ggplot2). Do not be afraid to experiment with R code you find online, modifying it to meet your needs (this is precisely why I have included code in this document and in the accompanying R script). I look up information nearly every time I use R, so I now have a list of helpful bookmarks. You will acquire a similar set of bookmarks as you become more familiar with R and/or modeling.

7. How do I know if I have fit my model correctly and/or that the model is correct?

   Mixed-effects modeling is complicated because it involves some trial and error, and modeling experts from different disciplinary traditions recommend different approaches. Moreover, as the field evolves, recommendations will change. All of these factors can make modeling intimidating, but do not feel intimidated! The best way to learn modeling, and to learn R, is to start modeling your own data with this guide and the other excellent introductions that have been published (Cunnings & Finlayson, 2015; Linck & Cunnings, 2015). As you model and write up results for publication, report your process and results as clearly as possible so that reviewers can offer assistance. Over time, you will become more confident and develop a more intuitive sense of how to fit and evaluate models. In short, learning how to model takes time.

8. What other resources can you recommend?

   First, know that you can always write me with questions related to modeling and I will do my best to answer them. In general, to help someone fit and evaluate models, it is helpful to have a description of the study and data set as well as the R code that is being used. You should also consult the references contained in this introduction. If you are interested in step-by-step guides, see Linck and Cunnings (2015) for a general overview, Cunnings and Finlayson (2015) for modeling longitudinal data, and Baayen (2008). The latter is very comprehensive and especially helpful for working with reaction time data but could be overwhelming for beginners. If you are interested in the theory behind mixed-effects models, especially as applied to longitudinal data, see Singer and Willett (2003). If you are interested in a general R statistics book that includes information on mixed-effects models, see Field, Miles, and Field (2013). Finally, there are many recent publications featuring mixed-effects models that can serve as excellent resources such as Barrios, Namyst, Lau, Feldman, and Idsardi (2016) and Offerman and Olson (2016).

ACKNOWLEDGMENTS

I would like to thank Nick Pandža (Faculty Research Specialist, University of Maryland Center for Advanced Study of Language) for first introducing me to mixed-effects modeling in R. Over the past five years, his expertise, encouragement, and friendship have proven invaluable. I would also like to thank Dr. Jared Linck (Research Scientist, University of Maryland Center for Advanced Study of Language) for his help during the early stages of modeling and Dr. Philip Dixon (University Professor of Statistics, Iowa State University) for helping me review and interpret models when I arrived at Iowa State. Finally, I am grateful to Dr. John Levis (Professor of TESL and Applied Linguistics, Iowa State University) and Dr. Pavel Trofimovich (Professor of Applied Linguistics, Concordia University) for encouraging me to present and publish on modeling and for supporting my broader research program. I am lucky to have had the help I have had over the years, and I hope I can provide similar assistance and encouragement to other researchers in the future.

ABOUT THE AUTHOR

Charles Nagle is an Assistant Professor of Spanish and the Director of the Spanish Language Program at Iowa State University. His main area of research is in second language pronunciation and individual differences. He is particularly interested in how learners' pronunciation develops over time, including the extent to which changing patterns of motivation and language use affect development, and the relationship between the perception and production of second language sounds. He is also passionate about statistics, particularly mixed-effects modeling. At Iowa State, he teaches all levels of Spanish language and upper-level linguistics courses such as "Introduction to Spanish Phonology" and "Bilingualism in the Spanish-Speaking World."

REFERENCES

Baayen, H. R. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. New York: Cambridge University Press.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. doi: 10.18637/jss.v067.i01

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255-278. doi:10.1016/j.jml.2012.11.001

Barrios, S. L., Namyst, A. M., Lau, E. F., Feldman, N. H., & Idsardi, W. J. (2016). Establishing new mappings between familiar phones: Neural and behavioral evidence for early automatic processing of nonnative contrasts. *Frontiers in Psychology, 7*, 1-16. doi:10.3389/fpsyg.2016.00995

Cunnings, I., & Finlayson, I. (2015). Mixed effects modeling and longitudinal data analysis. In L. Plonsky (Ed.), *Advancing Quantitative Methods in Second Language Research* (pp. 159-181). New York: Routledge.

Field, A., Miles, J., & Field, Z. (2013). *Discovering statistics using R*. Los Angeles: Sage.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. doi: 10.18637/jss.v082.i13

Linck, J. A., & Cunnings, I. (2015). The utility and application of mixed-effects models in second language research. *Language Learning, 65*(S1), 185-207. doi:10.1111/lang.12117

Murakami, A. (2016). Modeling systematicity and individuality in nonlinear second language development: The case of English grammatical morphemes. *Language Learning*, *66*(4), 834-871. doi:10.1111/lang.12166

Nagle, C. (2017a). Individual developmental trajectories in the L2 acquisition of Spanish spirantization. *Journal of Second Language Pronunciation, 3*(2), 219-242.

Nagle, C. (2017b). A longitudinal study of voice onset time development in L2 Spanish stops. *Applied Linguistics*. doi:10.1093/applin/amx011

Offerman, H. M., & Olson, D. J. (2016). Visual feedback and second language segmental production: The generalizability of pronunciation gains. *System, 59*, 45-60. doi:10.1016/j.system.2016.03.003

R Core Team. (2019). R: A language and environment for statistical computing (Version 3.5.3). Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis*. New York: Oxford University Press.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. New York: Springer.

Comeaux, K. C., & Thomson, R. I. (2019). Towards a deeper, uh, understanding of, um, L2 fluency and its [750 ms] silence. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Ames, IA, September 2018 (pp. 106-115). Ames, IA: Iowa State University.

PRESENTATION/POSTER

TOWARDS A DEEPER, UH, UNDERSTANDING OF, UM, L2 FLUENCY AND ITS [750 MS SILENCE] CORRELATES

Katie C. Comeaux, Brock University
Ron I. Thomson, Brock University

Previous research indicates that filled pauses (Fox Tree, 2001) and pauses that occur at clause boundaries (Brennan & Schober, 2001) tend to be less deleterious to listener judgments of intelligibility and comprehensibility than pauses produced clause-internally (Kang, 2010). Beyond their impact on ease of processing for listeners, hesitation phenomena may also lead to negative social evaluations of the speakers, particularly if pausing patterns are outside of listeners' linguistic- or culturally bound expectations. The current study aimed to confirm and extend these previous findings through a careful manipulation of natural L2 English speech stimuli, which were then presented to native speakers of English for evaluation. The L2 speech samples used were originally produced by 10 L1 Mandarin and 10 L1 Slavic talkers in the context of an extemporaneous picture description task. Each sample was then carefully manipulated to arrive at five matched versions that were either free of hesitation markers, included hesitation markers at clause boundaries (e.g., um, uh, or silence), or included hesitation markers placed within clauses. Using 9-point Likert-type scales, twenty listeners rated the speech samples for the speakers' fluency, comprehensibility, intelligence (IQ) and socio-economic status (SES). Findings suggest that unfilled pauses located at clause boundaries have a more positive impact on listeners' judgements of fluency, comprehensibility, intelligence, and SES.

INTRODUCTION

Hesitation phenomena (e.g., filled and silent pauses) are commonly observed in speech produced by both Native Speakers (NSs) and Non-native speakers (NNS). While the underlying cause of some hesitations evidences speakers' attempts to buy time during online planning and processing (i.e, cognitive disfluency), pausing also provides interactional cues to interlocutors. Thus, depending on the location and type of pause used, hesitations can have both facilitative and deleterious effects on listeners' comprehension (Fox Tree, 2001; Maclay & Osgood, 1959). Further, although the inclusion of pauses in spoken language may serve the same purpose for both NSs and NNSs, in second language (L2) speech, pauses often reflect real differences in cognitive fluency. L2 learners often have different knowledge of the target language, relative to NSs, and less efficient access to that knowledge. Consequently, listeners often judge NNS's use of pauses more harshly than those of NSs (Riggenbach, 1991), since hesitations may be taken to evidence an impoverished linguistic system. For example, pausing in second language (L2) speech is known to have harmful effects on the perception of a speaker's oral fluency and comprehensibility (Cenoz, 2000). Further, the way speakers utilize, and listeners perceive, pausing may be a culturally mediated act, meaning that certain pause types may be appropriate in one language context, but problematic in another (Cenoz, 2000; Watanabe, Hirose, Den, & Minematsu, 2008).

In this exploratory study, we build upon existing literature to further examine the extent to which (a) judgments of NNS's speech are influenced by the pause type utilized, and (b) whether a pause's location also impacts listener judgements of NNS's fluency, comprehensibility, intelligence and socio-economic status (SES), in a Canadian context.

Background

Despite several decades of research examining pausing phenomena, few studies have attempted to discriminate which pausing features are most detrimental to L2 communication, and of those studies, only a handful have utilized experimental methodologies (e.g., Bosker, Pinget, Quené, Sanders, & De Jong, 2013; Bosker, Quené, Sanders, & De Jong, 2014a; Bosker, Quené, Sanders, & De Jong, 2014b; Derwing, Munro, & Thomson, 2008; Kang, 2010; Kahng, 2018). What is clear is that the location and type of pause encountered in L2 speech matters to listeners.

Previous research reports strong agreement that pauses located at clause boundaries (CB) are less harmful to listener comprehension than pauses located clause internally (CI) (Lennon, 1990; Maclay & Osgood, 1959; Riggenbach, 1991; Watanabe et al. 2008). For example, it may be more difficult for listeners to process the sentence, "A man and woman bumped *uh* into each other and fell down", than the sentence "A man and a woman bumped into each other and *uh* fell down" Differences in the ease of processing such sentences are assumed to result from listeners' perceptions that pauses located at CB are expected, and less disruptive in comparison to pauses produced at CI locations (Maclay & Osgood, 1959; Riggenbach, 1991). Lennon (1990) reports that pauses located at CB's are perceived as being shorter than their CI counterparts of the same length. Additionally, it has been found that pauses located CI can create syntactic ambiguities (Watanabe et al., 2008). Bailey and Ferreira (2003) found that listeners struggle in understanding compound sentences when disfluencies follow head nouns (e.g., "The man *uh* went to the store and bought some booze"). With agreement across previous studies, it appears that pauses located at clause boundaries are less deleterious than pauses located clause internally.

Previous research has also found evidence to suggest that filled pauses differentially affect listeners' perception relative to unfilled pauses, although there is not complete agreement in these results (Blau, 1991; Brennan & Schober, 2001; Brennan & Williams, 1995; Clark & Fox Tree, 2002; Kang, 2010; Riggenbach, 1991; Watanabe et al., 2008). Kang (2010) found that unfilled pauses (UP) negatively influenced listener judgements of comprehensibility and accentedness, whereas filled pauses (FP) had no effect. In addition to listener judgements, FPs were found to lead to faster response times (Watanabe et al., 2008), and faster word recognition when utterances and words were following FPs (Brennan & Williams, 1995). Although there are some studies which suggest that UPs are less detrimental to listener processing than FPs (Cenoz, 2000; Clark & Fox Tree, 2002), the majority of the available evidence suggests that FPs may be less harmful to listener judgements than UPs.

While the present study seeks to further examine how pause location and types affect listener judgements of fluency and comprehensibility, it also examines the impact of pause location and pause type on social judgements of intelligence and SES. Previous research suggests that listeners are prone to making negative social judgements of foreign accented speech (Davila, Bohara, & Saenz, 1993; Munro, 2003; Munro & Derwing, 1995; Rubin, 1992). Davila et al. (1993) found that listeners perceived accented speakers as having lower incomes. Additionally, Rubin (1992) found that the lectures of university teachers who were perceived as having foreign accents were less understandable to students. Such social evaluations may, at least in part, be triggered by some measurable feature or features of the speech signal. As pausing phenomena vary across languages and cultures, the utilization of certain pausing strategies might then contribute to the perceived accentedness of a speech signal and may further contribute to negative social evaluations of NNSs.

In light of the previous research briefly described above and the questions raised therein, we propose two research questions to guide the current study.

Research questions

1. Do the locations of pauses (clause internal vs. at clause boundary) and the types of pauses (filled vs. unfilled) differentially affect listeners' perceptions of L2 fluency and comprehensibility? If so, which pause locations and types have stronger negative effects?
2. Do L2 English speakers' pause locations and pause types influence listeners' impression of the speakers' intelligence and socio-economic status?

METHODOLOGY

Participants

Twenty English NS listeners were recruited via conspicuously placed posters at a university in Southern Ontario. The listeners were primarily undergraduate students and ranged in age from 19-53 (mean age 25.4), and comprised eight males, eleven females and one non-binary person. Most of the participants were born and had resided in Southern Ontario for most of their lives. Fifteen of the participants reported having studied one or more second languages, however only two reported being fluent in a second language.

Materials

The speech samples utilized in this study were modified from samples used in Isaacs and Thomson (2013), where 20 beginner level, adult English as a second language (ESL) speakers (ten Mandarin L1; ten Russian L1) completed an eight-frame picture description task illustrating a humorous story about a man and a woman who mixed up their suitcases while traveling in a big city (Derwing et al., 2004). Isaacs and Thomson (2013) extracted the first 20 seconds of speech from each learner's description for analysis.

For the purposes of the current study, the 20-second L2 English speech samples used in Isaacs and Thomson (2013) were edited using Wave Pad software to synthesize five pausing conditions: 1) filled pauses at clause boundaries (FPCB); 2) filled pauses clause internally (FPCI); 3) unfilled pauses at clause boundaries (UPCB) 4) unfilled pauses located clause internally (UPCI); and 5) versions that are completely free of filled and unfilled pauses, that is hesitation free (HF). Versions with pauses (filled or unfilled) had between 1 and 3 pauses, depending on the number of clauses available for manipulation. Filled pauses were harvested from those produced by the same speaker within their unmodified speech samples and copied to the desired locations. Unfilled pauses were also based on each sample speaker's natural unmodified productions, with the same background noise as the rest of the recording, but moved when necessary to create the desired experimental conditions. Silent pauses were defined as any silence longer than 200 ms, a threshold previously identified as disruptive (Goldman-Eisler, 1961). Pause length in the UP conditions were not matched to FP counterparts, in an effort to maintain naturalness. It is possible that pause length could have a deleterious effect on listener judgements, however this requires further investigation in future studies. HF speech samples were modified to remove all filled pauses and all unfilled pauses over 200 ms. In total, these five manipulations resulted in 100 modified speech samples (see Table 1 for examples of each condition).

Table 1

Pausing conditions compared in this study

TYPE	ABBREVIATION	EXAMPLE
Filled pauses clause internal	(FPCI)	"Then they hit uh each other - their suitcase fell down"
Filled pauses at clause boundaries	(FPCB)	"Then they hit each other - uh - their suitcase fell down"
Unfilled pauses clause internal	(UPCI)	"Then they hit <silent pause> each other - their suitcase fell down"
Unfilled pauses at clause boundaries	(UPCB)	"Then they hit each other <silent pause> their suitcase fell down"
Hesitation free versions	(HF)	"Then they hit each other. Their suitcase fell down"

Procedure

Two questionnaires containing 9-point Likert-type scales were utilized to measure listener judgements of fluency, comprehensibility, intelligence, and SES as these scales are known to be reliable measures of listener perception (Derwing, Munro, & Thomson, 2008). The speech measures and the social measures were separated across two questionnaires with comprehensibility and intelligence paired, and fluency and SES paired to avoid the potential of cross-construct influence had we paired speech or social constructs together.

Rating sessions took place in a quiet room in the university. The definitions for each construct were provided prior to the rating sessions to ensure common understanding. Following these brief explanations, two practice items were played and rated to ensure the participants had the opportunity to ask questions and become comfortable with the task. The 100 randomized test items were played twice, once to rate comprehensibility and intelligence and once to rate fluency and

SES. Raters were given a five-minute break between rating sessions to reduce fatigue. The total time to conduct the experiment was one hour and twenty minutes.

Listener judgments were analyzed for interrater reliability using Chronbach's Alpha. Additionally, descriptive statistics and ANOVAs were conducted to identify any differences across pausing conditions, for fluency, comprehensibility, intelligence and SES. Correlations between pausing conditions for each of these constructs were also conducted.

RESULTS

Chronbach's alpha coefficients indicated strong interrater agreement on ratings for each construct: Fluency, .900; Comprehensibility, .952; Intelligence, .937; and Socio-economic status, .942. Thus, raters' scores were collapsed to examine mean ratings for each sample.

A repeated measures ANOVA provided a mixed result. Pauses appear to most impact Comprehensibility ratings, with significant effects on the ratings for both pause location [F (1, 19) = 5.379, p = 0.032, η2 = 0.221], and pause type [F (2, 38) = 6.138, p = 0.005, η2 = 0.244]. Raters preferred samples with pauses at clause boundaries, and preferred UPs over FPs. For Fluency, only pause type [F (2, 38) = 17.402, p < 0.000, η2 = 0.478] was found to have a significant effect on judgments. That is samples with FPs were perceived as significantly less fluent than samples with UPs. Similarly, for both IQ [F (2, 38) = 5.819, p = 0.006, η2 = 0.234] and SES [F (2, 38) = 4.526, p = 0.017, η2 = 0.192], again, pause types matter (raters prefer UPs) but not pause location. It is also worth noting that the ANOVA revealed no significant interactions between pause type and pause location within any construct. Effect sizes for pause type ranged from small to medium across constructs.

Descriptive statistics for the means (see Figure 1) illustrate these general patterns for all speech constructs (i.e., fluency, comprehensibility, IQ and SES). Hesitation free samples (i.e., those that had no filled nor silent pauses longer than 200 ms) were most preferred. Samples containing filled pauses were rated more negatively than samples containing unfilled pauses. Within the filled pause conditions, pauses located clause internally had either no impact or a more negative impact on ratings than pauses located at pause boundaries. Given the relatively small sample sizes, we do not feel a greater examination of variance patterns is warranted for this exploratory study.

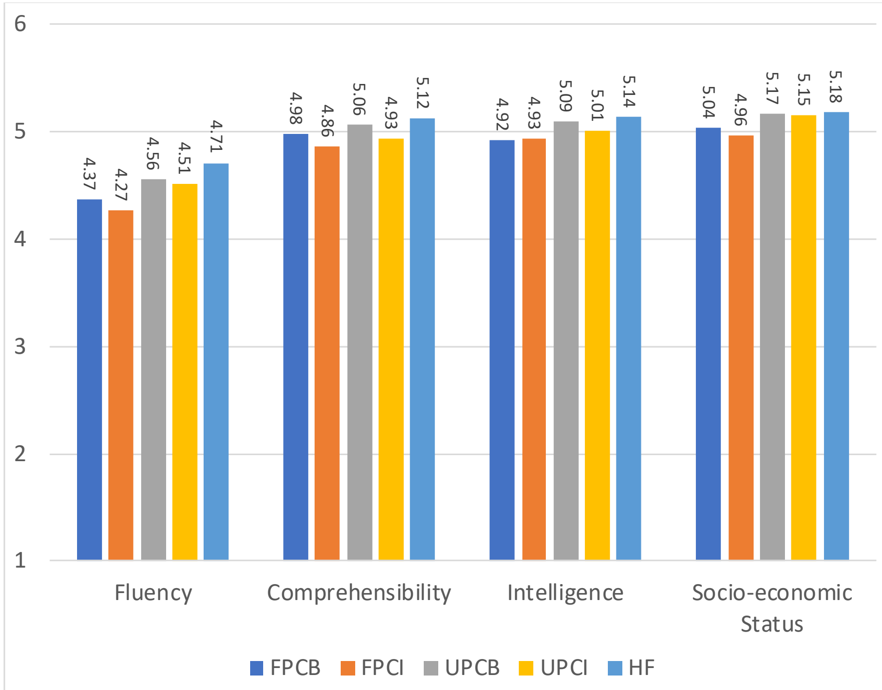


Figure 1. Rater means across pausing conditions (Filled pauses at clause boundaries (FPCB); Filled pauses clause internally (FPCI); Unfilled Pauses at clause boundaries (UPCB); Unfilled pauses clause internally (UPCI); and Hesitation free (HF) speech.

We also examined the extent to which individual raters, on average, preferred speech samples with UP over FP. In 70% of cases the unfilled pause version of a sample was preferred (i.e., rated more favorably). In 25% of cases the filled pause version was preferred. And in 5% of the cases there was no preference. This demonstrates that raters were largely consistent with each other and were affected by the same speech features in similar ways.

Across all constructs and pause types, there was a difference in ratings across speakers. For instance, the Mandarin speakers consistently received lower ratings than the Russian speakers in every construct and pausing condition. This is likely not meaningful but reflects the fact that the Mandarin speakers, as a group, were less proficient in English and therefore less fluent. While it might also be possible that the nature of the filled pauses used by the Mandarin vs. the Slavic speakers impacts ratings, and this should not be ruled out, the speech rate of the Mandarin speakers was slower (see Isaacs and Thomson, 2013), and so the effect of L1 cannot be disentangled from their overall temporal fluency.

Pearson correlation coefficients were also conducted between the speech measures (mean fluency and comprehensibility) and the social judgments (mean IQ and SES) (See Table 2). Correlations between fluency and the social dimensions were extremely high in both filled pause and unfilled pause contexts. Similarly, correlations between comprehensibility and the social dimensions were extremely high. It is worth recalling that fluency and IQ were not rated simultaneously, nor were comprehensibility and SES. Therefore, the strength of these correlations cannot be attributed to the influence of rating these constructs at the same time.

Table 2

Pearson correlations between mean speech measures and social measures

	IQ Filled pause condition	SES Filled pause condition
Fluency Filled pause condition	.952**	.963**
Comprehensibility Filled pause condition	.958**	.925**
	IQ Unfilled pause condition	SES Unfilled pause condition
Fluency Unfilled pause condition	.992**	.973**
Comprehensibility Unfilled pause condition	.965**	.913**

DISCUSSION

The current study found that pause type significantly impacts listener judgements of fluency, comprehensibility, intelligence and SES of L2 speakers. Filled pauses, regardless of their location, had a significant negative effect on judgements of fluency and comprehensibility. This is an interesting finding, as some previous research suggests that unfilled pauses may be more deleterious to the listener. One possible explanation for this finding could be the length of pauses, which were, in an effort to maintain naturalness of speech samples, not strictly controlled for in this study. Had pause durations for FP and UP conditions been identical, a different result may have been obtained. For example, a long silent pause broken up by a filled pause may be preferable to a long silent pause without a filler breaking it up. Since we shortened the length of overly long silent pauses after removing fillers, we cannot answer this question. Future research should consider investigating the duration threshold at which silent pauses may become more problematic for listeners than the same amount of silence interrupted by filled pauses.

The location of pauses seemed to matter less than pause type, with a significant negative effect for comprehensibility only, when either filled or unfilled pauses were heard within clauses. The lack of significance for pause location, outside of the comprehensibility construct, may be due to the fact that in order to judge for comprehensibility, listeners need to attend to the meaning of a message. The other measures may only require that a listener attend to some other details (e.g., speech rate) to form judgements on the speech samples. Although our study does support previous findings that pauses located at clause boundaries are less deleterious than pauses located clause internally, it appears that depending on the purpose of listening, it may not matter where a pause is located, as long as the pause type is facilitative.

NS listeners willingly made social judgements on NNS speech and were affected by pausing conditions in a manner that reflected responses in terms of the speech measures (fluency and comprehensibility). Both IQ and SES ratings were negatively affected by filled pauses, regardless of location. Perceived fluency and comprehensibility of speech samples were also very strongly correlated with judgements of the speaker's IQ and SES. This seems to suggest that judgments about these social dimensions are really just based on the perceived fluency of the speaker, which may be a reasonable if faulty basis for such judgments. Whatever the explanation, evidence from this study supports previous research, which indicates that NS listeners can form negative judgments of L2 speakers on the basis of features in their speech. The hesitation free samples which would, apart from accentedness, most closely represent NS speech, were the most highly rated for both speech and social measures. As pausing may be more often utilized in NNS' speech (especially those of lower proficiency), the likelihood of negative listener evaluation increases. This study is limited in its ability to determine whether NS listeners are affected by pause type and location in similar ways when NS samples are played, but in future studies, NS controls could help determine the extent to which these negative judgements are due to the accented speech itself, or the features found in accented speech, such as pause type and pause location. It would also be beneficial to explore whether pauses located at different clause boundaries (e.g. NP, VP) are perceived similarly. Since some raters expressed feeling uncomfortable making social judgements, using a follow up questionnaire may have been beneficial to understanding why participants made the social judgements that they did, especially considering that despite the reported discomfort, there was still a clearly reliable relationship between pausing phenomena and social evaluation.

CONCLUSION

The findings of this exploratory study both support and contradict previous research. This suggests that further research is necessary to better understand the complexity in the use of pausing strategies by NNSs. Conducting a more controlled experiment where speech samples are further manipulated to determine whether pause duration impacts listener judgments, and where native speaker samples are included will lead to more conclusive results. It would also be beneficial to examine the different effects of lexical (e.g. like) versus non-lexical filled pauses since the current study was only limited to non-lexical filled pauses.

Though the results require further investigation, there may be some preliminary application for our findings. Since unfilled pauses appear to be less detrimental to listener judgements of fluency, comprehensibility, intelligence and socio-economic status, there may be some value in encouraging L2 learners to use brief silent pauses rather than filled pauses during online planning for speech production. The importance of teaching pausal phenomena extends beyond direct communication and may lead to adverse social judgements on the speakers themselves. L2 learners may be empowered if given the tools to communicate in ways that are accepted by their L2 community, while also raising their awareness that their interlocutors may make arbitrary social judgements, based on the type of pausing strategies they use. Additionally, the wider community of listeners may also benefit from knowledge that they may be making negative social judgements on the basis of features of the L2 speech signal from which the social and personal attributes of the speaker cannot actually be deduced (Munro, 2003).

REFERENCES

Bailey, K. G. D., & Ferreira, F. (2003). Disfluencies affect the parsing of garden-path sentences. *Journal of Memory and Language*, *49*, 183-200.

Blau, E. K. (1991). *More on comprehensible input: The effect of pauses and hesitation markers on listening comprehension.* Paper presented at the Annual Meeting of the Puerto Rico Teachers of English to Speakers of Other Languages, San Juan, Puerto Rico. Retrieved from https://files.eric.ed.gov/fulltext/ED340234.pdf.

Bosker, H. R., Pinget, A. F., Quené, H., Sanders, T., & De Jong, N. H. (2013). What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing*, *30*(2), 159-175.

Bosker, H. R., Quené, H., Sanders, T., & De Jong, N. H. (2014a). The perception of fluency in native and nonnative speech. *Language Learning*, *64*(3), 579-614.

Bosker, H. R., Quené, H., Sanders, T., & de Jong, N. H. (2014b). Native 'um's elicit prediction of low-frequency referents, but non-native 'um's do not. *Journal of Memory and Language*, *75*, 104-116.

Brennan, S. E., & Schober, M. F. (2001). How listeners compensate for disfluencies in spontaneous speech. *Journal of Memory and Language*, *44*, 274-296.

Brennan, S. E., & Williams, M. (1995). The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory and Language*, *34*, 383-398.

Cenoz, J. (2000). Pauses and hesitation phenomena in second language production. *International Journal of Applied Linguistics*, *127*(1), 53-69.

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*, *84*, 73-111. doi:10.1016/S0010-0277(02)00017-3.

Davila, A., Bohara, A. K., & Saenz, R. (1993). Accent penalties and the earnings of Mexican Americans. *Social Science Quarterly*, *74*(4), 902-916.

Derwing, T. M., Munro, M. J., & Thomson, R. I. (2008). A longitudinal study of ESL learners' fluency and comprehensibility development. *Applied Linguistics*, *29*, 359-380.

Fox Tree, J. E. (2001). Listeners' uses of um and uh in speech comprehension. *Memory and Cognition*, *29*(2), 320-326.

Goldman-Eisler, F. (1961). A comparative study of two hesitation phenomena. *Language & Speech*, *4*(1), 18-26.

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.

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

Kahng, J. (2018). The effect of pause location on perceived fluency. *Applied Psycholinguistics*, *39*(3), 569-591.

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, *40*(3), 387-417.

Maclay, H., & Osgood, C. (1959). Hesitation phenomena in spontaneous English speech. *WORD*, *15*(1), 19-44.

Munro, M. J. (2003). A primer on accent discrimination in the Canadian context. *TESL Canada Journal*, *20*(2), 38-51.

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, *45,* 73-97. doi:10.1111/j.1467-1770.1995.tb00963.

Riggenbach, H. (1991). Toward understanding fluency: A micro-analysis of nonnative speaker conversation. *Discourse Processes*, *14*, 423-441.

Rubin, D. L. (1992). Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher Education*, *33*(4), 511.

Skehan, P., Foster, P., & Shum, S. (2016). Ladders and snakes in second language fluency. *International Review of Applied Linguistics in Language Teaching*, *54*, 97-111. doi:10.1515/iral-2016-9992.

Watanabe, M., Hirose, K., Den, Y., & Minematsu, N. (2008). Filled pauses as cues to the complexity of upcoming phrases for native and non-native listeners. *Speech Communication*, *50*, 81-94. doi:10.1016/j.specom.2007.06.002.

Gordon, J., & Darcy, I. (2019). Teaching segmentals vs. suprasegmentals: Different effects of explicit instruction on comprehensibility. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Ames, IA, September 2018 (pp. 116-126). Ames, IA: Iowa State University.

PRESENTATION/POSTER

TEACHING SEGMENTALS VS. SUPRASEGMENTALS: DIFFERENT EFFECTS OF EXPLICIT INSTRUCTION ON COMPREHENSIBILITY

Joshua Gordon, University of Northern Iowa
Isabelle Darcy, Indiana University

This investigation reports the results of a pronunciation intervention to enhance the comprehensibility of learners of English as a foreign language (EFL) at a university in Costa Rica. Three groups of EFL students received pronunciation instruction on segmentals, suprasegmentals, or a combination of both using explicit phonetic instruction and communicative tasks during 10 weeks (30-mins each week). We collected spontaneous speech samples in a pretest and a posttest, and presented them to a group of 40 native speakers of English to be rated for comprehensibility. Our results indicated that the group trained in suprasegmentals significantly improved in comprehensibility by the end of the intervention. Additionally, the group of learners who received instruction with a combination of segmentals and suprasegmentals also improved in comprehensibility at the end of the intervention—although not significantly. Finally, the group trained on segmentals did not improve in comprehensibility. Our findings suggest that an intervention on suprasegmentals seems to help learners in the development of comprehensibility in a short period of time (see Gordon & Darcy, 2016; Levis & Muller Levis, 2018), and that incorporating pronunciation little by little into the regular language classroom can help learners achieve comprehensible speech in the long run (Darcy, 2018; Derwing, Munro, & Wiebe, 1998; Sardegna, Chiang, & Ghosh, 2016; Sicola & Darcy, 2015).

INTRODUCTION

The field of second language (L2) pronunciation teaching and research is currently experiencing renewed interest, which is evident in an increase in the number of studies in recent years (Thomson & Derwing, 2015). In spite of growing research evidence, new findings are not necessarily incorporated in L2 classrooms due to factors such as limitations in teacher training (Baker, 2014; Foote, Holtby & Derwing, 2011; Huensch, 2018; Murphy, 2014), or the lack of interest in pedagogical interventions on the part of researchers (Derwing & Munro, 2015). Nevertheless, more collaboration between researchers and teachers is crucial for implementation of research-based classroom practices that could help L2 learners enhance their pronunciation (Derwing & Munro, 2005, 2015; Levis, 2016). In this study, we present the results of a classroom pronunciation intervention using an experimental treatment anchored in research evidence that aimed at improving the comprehensibility of a group of English-as-a-foreign-language (EFL) learners in Costa Rica. The following literature review presents the theoretical background that motivated our study.

LITERATURE REVIEW

Recent meta-analyses that investigated the value of pronunciation instruction (e.g., Lee, Jang, &

Plonsky, 2015; Saito, 2012) have confirmed its effectiveness to enhance intelligible (i.e., actually understandable) and comprehensible (i.e., easy to understand) L2 speech (see Derwing & Munro, 2009). Effectiveness of instruction is indexed through improvement in phonetic accuracy (accent change), or through speech that is rated as more comprehensible (see Thomson & Derwing, 2015). Additionally, research has demonstrated that pronunciation instruction can be effective not only in beginner learners (see Zielinski & Yates, 2014), but also in L2 learners who have spoken the language for a long time in different settings and whose pronunciation presents entrenched characteristics (e.g., Derwing, Munro, & Wiebe, 1997; Derwing, Munro, Foote, Waugh, & Fleming, 2014).

Importantly, the research scope of most pronunciation studies carried out in recent years has been consistent. For example, in their meta-analysis on the efficacy of pronunciation instruction, Lee et al. (2015) found that the majority of pronunciation studies administered in recent years have focused on the development of segmentals (e.g., Baker & Trofimovich, 2006; Kissling, 2013; Saito, 2013a, 2013b), whereas fewer studies have investigated the development of suprasegmentals (Derwing, Rossiter, Munro, & Thomson, 2004; Hahn, 2004; Saito & Wu, 2014; Trofimovich & Baker, 2006), or a combination of both (Derwing, Munro, & Wiebe, 1998; Derwing & Rossiter, 2003; Saito & Saito, 2016). In a similar manner, Thomson and Derwing (2015) mentioned that most of the studies in their narrative review of 75 pronunciation studies (63%) were guided by accent or nativeness principles, as opposed to the 24% of those studies that were guided by intelligibility principles or the 13% that included a combination of both (see Levis, 2005, for a discussion of these principles).

Because the research evidence indicates that explicit pronunciation instruction can be effective and that it can help learners in the development of comprehensible speech (Thomson & Derwing, 2015), we implemented a pedagogical intervention in this study to enhance the comprehensibility of EFL learners. The methodological steps we applied are presented below.

THE CURRENT STUDY

In this pronunciation intervention, we investigated the effects of implementing three types of explicit pronunciation instruction in three groups of EFL learners. The study was guided by two research questions:

1. Do EFL students improve their comprehensibility by the end of a 10-week pronunciation intervention in the classroom?
2. If so, which type of explicit instruction (based on segmentals, suprasegmentals, or both) leads to more comprehensible speech?

Participants

Speakers. Three intact EFL classes at a small university in Costa Rica received pronunciation instruction during a period of 10 weeks for 30 minutes each week. We collected speech samples from students in these classes in a pretest and posttest in the form of video-description narratives. All the EFL learners were enrolled in a first semester of English class, and their ages ranged from 17 to 21[i]. They were undergraduate students of Computer Engineering and Tourism, and their L1 was Spanish. Each class was composed of about 25 students; however, because of logistics constraints common in this type of research (e.g., student absences during treatment), we considered only students who completed all the training sessions (see Derwing & Munro, 2015; Mackey & Gass, 2016). Therefore, a total of 22 students were included in the final analyses (i.e., 7 students from the Segmental group, 8 students from the Suprasegmental group, and 7 students from the Mixed group). Additionally, two English as a first language (L1) speakers provided speech samples to be used as a baseline. They were male and female, ages 19 and 21 respectively, both undergraduate students at an American university in the Midwest. These participants recorded the speech samples only once**.**

Raters. A group of 40 English L1 speakers rated the speech samples from the group of EFL learners (and the two English L1 speakers) for comprehensibility – on a scale of 1 to 9 where 1=*extremely easy to understand* and 9=*impossible to understand* (see Munro & Derwing, 1995). The raters were undergraduate students at an American university in the Midwest, between 17 and 21 years of age. All were from the same Midwestern region for the purpose of uniformity. They were all untrained raters, and did not speak any language fluently other than English. They were enrolled in a second semester Spanish language class at the time of the study.

Treatment

This study followed a pretest-posttest design; the three groups of EFL learners received treatment based on segmentals, suprasegmentals, or a combination of both over 10 weeks for 30 minutes each week. As for the sequence of activities in the treatments, we used Celce-Murcia, Brinton, Goodwin, and Griner's (2010) communicative framework for pronunciation instruction, which included activities in a continuum from explicit phonetic instruction to communicative meaning-based activities. Table 1 below summarizes the sequence of different activities followed by each group.

Table 1

Sequential treatment in three experimental groups

Instructional Stages and Techniques	**Segmental Group**	**Suprasegmental Group**	**Mixed Group**
Description & Analysis / Listening Discrimination	Explicit instruction and analysis of segmental features: *-Oral introduction of topic* *-Visual aids* *-Listening discrimination tasks*	Explicit instruction and analysis of suprasegmental features: *-Oral introduction of topic* *-Visual aids* *-Listening discrimination tasks*	Explicit instruction and analysis of segmental and suprasegmental features: *-Oral introduction of topic* *-Visual aids* *-Listening discrimination tasks*
Controlled & Guided Practice	*-Minimal pair drills* *-Analyses of words and phrases* *-Reading short passages*	*-Drills (word and sentence stress)* *-Analyses of words and phrases* *-Reading short passages*	*-Minimal pair drills* *-Drills (contrastive stress)* *-Analyses of words and phrases* *-Reading short passages*
Communicative Practice	Meaning-oriented activities *-Task-based communicative activities*	Meaning-oriented activities *-Task-based communicative activities*	Meaning-oriented activities *-Task-based communicative activities*

As for the selection of content, the Segmental group received instruction on vowels [i, ɪ, ɛ, æ, ʌ, ɑ, u, ʊ] and consonants [p, t, k, b, d, g], which were selected based on functional load criteria—that is, segments with a high functional load when they occur in different minimal pairs (Brown, 1991; Munro & Derwing, 2006). The Suprasegmental group, in contrast, received instruction on prosodic aspects of English such as word and sentence stress, rhythm, intonation as well as aspects like linking, contractions, and vowel reduction. Finally, the Mixed group received instruction on both segmentals and suprasegmentals similar to the activities designed for the other groups.

Pretest and posttest

We collected speech samples from the learners in the form of a pretest (at Time 1 or at the beginning of their course) and a posttest (at Time 2, or during week 11 after treatment). The participants recorded descriptions of two short video cartoons found on the internet (*Simon's Cat*, 2009, 2010) at Times 1 and 2 (a different video at each time; both were about 2:30 minutes long). We asked the participants to watch each video completely and to pay attention to the story depicted. After that, we asked them to describe what happened in each video giving as many details as possible. The recordings were made individually on a personal computer (13-inch Macbook Pro, with a Logitech USB Headset H390) in a quiet room at a library, and we used the speech software Praat (Boersma & Weenink, 2016, version 6.0.15) to record each participant. There was a specific action in the plot of each video that was described by all the participants, and we selected 20 seconds of the description of those actions to be presented to the group of English as an L1 raters for uniformity purposes. As noted above, two English L1 speakers recorded descriptions of both videos only once.

Rating task

The rating task was conducted in a computer lab using high-quality headphones. Before completing the task, the raters watched the two video cartoons described by the speakers to avoid biased ratings with the first speech samples (see Derwing et al., 2004). They also completed a short warm-up in which they rated five speech samples (produced by speakers from another study) to get familiar with the task. We presented the speech samples from all the speakers to the group of 40 raters using the survey software Qualtrics. They heard the randomized samples from the pretest and posttest, which they rated for comprehensibility using a Likert scale from 1 to 9, where 1=*extremely easy to understand* and 9=*impossible to understand* (Munro & Derwing, 1995). The results of the task are presented below.

DATA ANALYSIS AND RESULTS

We carried out a linear-mixed effects model using Comprehensibility Ratings as dependent variable, declaring Group (Segmental, Suprasegmental, or Mixed) and Time (pretest & posttest) as fixed effects, and Speaker and Rater as random effects. There was a significant effect of Group ($F(3, 14) = 8.55$, $p = 0.0018$), which confirmed that the EFL learners and the L1-English speakers were judged very differently by the raters. As expected, the L1 baseline speakers were the most comprehensible speaker group. Table 2 below shows the mean scores obtained by each group during the pretest and posttest.

Table 2

Mean comprehensibility scores of L2-learner groups and native speakers

Group	***M* Time 1**	**SD**	***M* Time 2**	**SD**
Segmental	3.92	1.73	4.35	1.87
Suprasegmental	4.58	1.92	4.21	1.93
Mixed	4.18	1.83	4.05	1.67
Native Speaker	1.21	0.91	1.18	0.65

In order to isolate results from the L2 learners, we carried out subsequent analyses with the three groups of EFL learners excluding the L1 speakers. Even without the baseline group, we found a significant *Group* by *Time* Interaction ($F(2, 1222) = 8.77$, $p < 0.001$), in which the Suprasegmental group significantly improved in comprehensibility from time 1 to time 2. The Mixed group also became more comprehensible from time 1 to time 2, but this increase in comprehensibility was not significant. Finally, the Segmental group also presented a significant difference from time 1 to time 2, but this difference resulted in a decrease in comprehensibility. Table 3 presents the main differences from pretest to posttest in each one of the groups. Figure 1 summarizes the results in comprehensibility in both the pretest and posttest in the three EFL groups.

Table 3

Mean difference from pretest to posttest in three L2-learner groups

Group	**Mean diff. T1 to T2**	**S.E.**	***t*-value**	***p* value**
Segmental	0.4350	0.1443	3.02	0.0026
Suprasegmental	-0.3708	0.1317	-2.82	0.0049
Mixed	-0.1300	0.1443	-0.90	0.3677

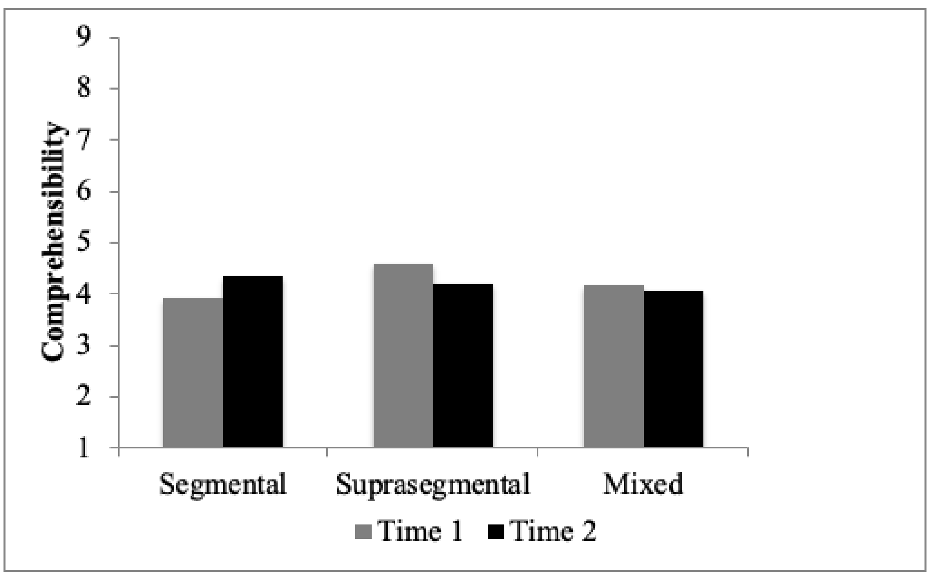


Figure 1. Comprehensibility ratings across time in three learner groups.

DISCUSSION AND CONCLUSIONS

Although this study presented common difficulties that are typical of this type of research (e.g., students with different proficiency levels, attrition, see Derwing & Munro, 2015), its modest results provide more support for the importance of explicit phonetic instruction in pronunciation teaching and learning. Our first research question asked, *Do EFL students improve their comprehensibility as a result of a 10-week pronunciation classroom intervention?* According to the results obtained, there were different types of improvement in both the Suprasegmental and

Mixed groups. Whereas the Suprasegmental group became significantly more comprehensible by the end of the experimental treatment, the improvement in comprehensibility shown by the Mixed group in the posttest was not significant. While this improvement in the Mixed group is certainly positive, it is possible that a treatment of 10 weeks of only 30 minutes a week is simply not enough to develop substantially more comprehensible L2 speech based on instruction of both segmentals and suprasegmentals at the same time in an EFL context. It is necessary to stress that whereas the Mixed and Suprasegmental groups improved their comprehensibility, the Segmental group was rated as less comprehensible by the end of treatment. It is important to remember that the three groups followed the same sequence of activities during treatment, but with different phonetic content. Thus, it is possible that the learners in the Segmental group mainly focused on producing segments accurately in spontaneous speech without using other aspects that are generally known to make speech more comprehensible (e.g., discourse organization, syntactic, lexical, and phonological accuracy). Derwing et al. (1998) reported a similar finding in their study in which a group trained on segmentals did not improve comprehensibility in spontaneous speech, possibly because of allocation of attentional resources. One of the possible interpretations Derwing and colleagues cite for their finding is that, because of their experimental training, the learners in their segmental group focused their attention on the production of segments. This in turn did not leave learners any attentional resources to use in a more demanding task like a spontaneous picture description (e.g., using grammatically-correct sentences, lexical retrieval, discourse organization, phonological accuracy). In a similar manner, it is possible that our learners in the segmental group also focused their attention on accuracy of segmental production, at the detriment of other aspects that are also necessary to develop comprehensibility in spontaneous speech (e.g., fluency, lexical and sentence stress, rhythm, appropriate pauses, intonation; see Isaacs & Trofimovich, 2012).

In the second research question, we asked, *Which type of explicit instruction (based on segmentals, suprasegmentals, or both) leads to more comprehensible speech?* Our results suggest that a treatment based on suprasegmentals seems most effective in a short period of time, at least in this particular context. The significance of this result is twofold: first, our results align with previous studies that demonstrated that a treatment on suprasegmentals during a short period of time can enhance comprehensibility (Gordon, Darcy, & Ewert, 2013; Gordon & Darcy, 2016; Levis & Muller Levis, 2018); secondly, these results again confirm the prominent role of suprasegmentals in the perception of comprehensible speech (Anderson-Hsieh, et al., 1992; Field, 2005, 2008; Hahn, 2004; Isaacs & Trofimovich, 2012; Kang, Rubin, & Pickering, 2010). This does not mean that there should not be any focus on segmentals in pronunciation instruction. In fact, it is widely recognized in the pronunciation field that segmentals are important for intelligibility at the lexical level and that any pronunciation intervention should consider a combination or “cocktail” of both segments and prosody (see Derwing et al., 1998; Zielinski, 2006). Additionally, previous studies have demonstrated that learners’ knowledge of segmentals can help disentangle confusion in the perception and production of minimal pairs (Derwing, Munro, & Wiebe, 1997, 1998). However, the results obtained here also confirm that training L2 learners in suprasegmentals can help them sound more natural and comprehensible in spite of the presence of a foreign accent (Derwing & Munro, 2015).

As a final note, these results confirm that embedding pronunciation instruction in speaking classes—even for short periods of time in each lesson—can help learners achieve comprehensible speech in the long run (Darcy, 2018; Derwing, Munro, & Wiebe, 1998; Sardegna, Chiang, &

Ghosh, 2016; Sicola & Darcy, 2015). These results also suggest once again that pronunciation instruction is likely to be effective when it makes use of a communicative component where learners practice the language in a continuum of activities that range from explicit instruction and controlled activities to more meaning-based communicative tasks (Celce-Murcia et al., 2010; Levis & Grant, 2003; Trofimovich & Gatbonton, 2006; Zielinski & Yates, 2014). Therefore the combination of explicit phonetic instruction with controlled and meaning-based activities could give learners the opportunity to put into practice—in both controlled tasks and spontaneous speech—the phonetic forms that they learn under more controlled conditions, in order to help them develop comprehensible and intelligible L2 speech.

REFERENCES

Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviances in segmentals, prosody, and syllable structure. *Language Learning*, *42*, 529-555.

Baker, A. (2014). Exploring teachers' knowledge of second language pronunciation techniques: Teacher cognitions, observed classroom practices, and student perceptions. *TESOL Quarterly, 48,* 136-163.

Baker, W., & Trofimovich, P. (2006). Perceptual paths to accurate production of L2 vowels: The role of individual differences. *International Review of Applied Linguistics*, *44*, 231-250.

Boersma, P., & Weenink, D. (2016). Praat: Doing phonetics by computer [Computer program] Version 6.0.15. Retrieved 23 March 2012 from http://www.praat.org/.

Brown, A. (1991). *Pronunciation models*. Singapore: Singapore University Press.

Celce-Murcia, M., Brinton, D., Goodwin, J., & Griner, B. (2010). *Teaching pronunciation: A reference for teachers of English to speakers of other languages* (2nd ed.). Cambridge, UK: Cambridge University Press.

Darcy, I. (2018). Powerful and effective pronunciation instruction: How can we achieve it? *The CATESOL Journal, 30,* 13-45.

Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly, 39,* 379-397.

Derwing, T. M., & Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching*, *42*, 476-490.

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

Derwing, T. M., Munro, M. J., Foote, J. A., Waugh, E., & Fleming, J. (2014). Opening the window on comprehensible pronunciation after 19 years: A workplace training study. *Language Learning*, *64*, 526-548.

Derwing, T., Munro, M., & Wiebe, G. (1997). Pronunciation instruction for fossilized learners: Can it help? *Applied Language Learning, 8,* 217-235.

Derwing, T., Munro, M., & Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning, 48,* 393-410.

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,* 1-17.

Derwing, T. M., Rossiter, M. J., Munro, M. J., & Thomson, R. I. (2004). Second language fluency: Judgments on different tasks. *Language Learning, 54,* 655-679.

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

Field, J. (2008). Bricks or mortar: Which parts of the input does a second language listener rely on? *TESOL Quarterly, 42,* 411-432.

Foote, J., Holtby, A., & Derwing, T. (2011). Survey of the teaching of pronunciation in adult ESL programs in Canada, 2010. *TESL Canada Journal, 29,* 1-22.

Gordon, J., & Darcy, I. (2016). The development of comprehensible speech in L2 learners: A classroom study on the effects of short-term pronunciation instruction. *Journal of Second Language Pronunciation, 2,* 56-92.

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,* 201-223.

Huensch, A. (2018). Pronunciation in foreign language classrooms: Instructors' training, classroom practices, and beliefs. *Language Teaching Research*, 1-20. https://doi.org/10.1177%2F1362168818767182

Kang, O., Rubin, D., & Pickering, L. (2010). Suprasegmental measures of accentedness and judgments of language learner proficiency in oral English. *Modern Language Journal, 94,* 554-566.

Kissling, E. M. (2013). Teaching pronunciation: Is explicit phonetic instruction beneficial for FL learners? *Modern Language Journal, 97,* 720-744.

Isaacs, T., & Trofimovich, P. (2012). Deconstructing comprehensibility: Identifying the linguistic influences on listeners' L2 comprehensibility ratings. *Studies in Second Language Acquisition, 34,* 475-505.

Lee, J., Jang, J., & Plonsky, L. (2015). The effectiveness of second language pronunciation instruction: A meta-analysis. *Applied Linguistics, 36,* 345-366.

Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly, 39,* 369-377.

Levis, J. M. (2016). Research into practice: How research appears in pronunciation teaching materials. *Language Teaching, 49,* 423-437.

Levis, J. & Grant, L. (2003). Integrating pronunciation into ESL/EFL classrooms. *TESOL Journal*, *12*, 13-19.

Levis, J. M., & Muller Levis, G. (2018). Teaching high-value pronunciation features: Contrastive stress for intermediate learners. *The CATESOL Journal, 30,* 139-160.

Mackey, A., & Gass, S. (2016). *Second language research: Methodology and design* (2nd ed.). New York, NY: Routledge.

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

Munro, M. J., & Derwing, T. M. (2006). The functional load principle in ESL pronunciation instruction: An exploratory study. *System, 34,* 520-531.

Murphy, J. (2014). Myth: ESL teachers are ready to teach pronunciation. In L. Grant (Ed.), *Pronunciation myths*. Ann Arbor: University of Michigan Press.

Saito, K. (2012). Effects of instruction on L2 pronunciation development: A synthesis of 15 quasi-experimental intervention studies. *TESOL Quarterly, 46,* 842-854.

Saito, K. (2013a). Reexamining effects of form-focused instruction on L2 pronunciation development: The role of explicit phonetic information. *Studies in Second Language Acquisition*, *35,* 1-29.

Saito, K. (2013b). The acquisitional value of recasts in instructed second language speech learning: Teaching the perception and production of English /ɹ/ to adult Japanese learners. *Language Learning, 63,* 499-529.

Saito, Y., & Saito, K. (2016). Differential effects of instruction on the development of comprehensibility, word stress, rhythm, and intonation: The case of inexperienced Japanese EFL learners. *Language Teaching Research*, *20,* 1-20.

Saito, K., & Wu, X. (2014). Communicative focus on form and second language suprasegmental learning: Teaching Cantonese learners to perceive Mandarin tones. *Studies in Second Language Acquisition, 36*, 647-680.

Sardegna, V., Chiang, F.-H., & Ghosh, M. (2016). Integrating pronunciation with presentation skills. In T. Jones (Ed.), *Pronunciation in the classroom: The overlooked essentials* (pp. 43-56). Alexandria, VA: TESOL Press.

Sicola, L., & Darcy, I. (2015). Integrating pronunciation into the language classroom. In M. Reed & J. Levis (Eds.), *Handbook of English pronunciation* (pp. 467-483). Malden, MA: Wiley.

Simon's Cat. (2009, July 24). *Fly guy* [Video file]. Retrieved from https://www.youtube.com/watch?v=I1qHVVbYG8Y

Simon's Cat. (2010, February 10). *Snow business* [Video file]. Retrieved from https://www.youtube.com/watch?v=Tuf61OjvoPQ

Thomson, R., & Derwing, T. (2015). The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics, 36,* 326-344.

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

Trofimovich, P., & Gatbonton, E. (2006). Repetition and focus on form in processing L2 Spanish words: Implications for pronunciation instruction. *Modern Language Journal*, *90*, 519-535.

Zielinski, B. (2006). The intelligibility cocktail: An interaction between speaker and listener ingredients. *Prospect*, *21*, 22-45.

Zielinski, B., & Yates, L. (2014). Myth: Pronunciation instruction is not appropriate for beginning-level learners. In L. Grant (Ed.), *Pronunciation myths: Applying second language research to classroom teaching* (pp. 56-79). Ann Arbor: University of Michigan Press.

[i] There was no specific parameter to determine the proficiency level of these students. Although there were different proficiency levels among the students due to previous instruction in secondary schools, most college students who come from public secondary schools in Costa Rica present a very basic or false-beginner proficiency level when they take their first EFL class at the university level.

Grenon, I., Sheppard, C., & Archibald, J. (2019). The effect of discrimination training on Japanese listeners' perception of the English coda consonants as in 'rose' and 'roads'. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Ames, IA, September 2018 (pp. 127-136). Ames, IA: Iowa State University.

PRESENTATION/POSTER

THE EFFECT OF DISCRIMINATION TRAINING ON JAPANESE LISTENERS' PERCEPTION OF THE ENGLISH CODA CONSONANTS AS IN 'ROSE' AND 'ROADS'

Izabelle Grenon, The University of Tokyo, Japan
Chris Sheppard, Waseda University, Japan
John Archibald, The University of Victoria, Canada

The efficiency of phonetic training via discrimination tasks has been questioned, as phoneme-grapheme correspondence is not transparent in discrimination training. Indeed, we showed in a previous study that some Japanese learners of English associated the vowels in 'ship' and 'sheep' with the wrong orthographic representation. The current study evaluated if mislabeling issues would occur when Japanese learners of English train with the contrast as in 'rose' and 'roads', and whether any improvement over time would be observed. Forty native English speakers from North America participated as the group of reference. Twenty Japanese speakers received two discrimination training sessions of about thirty minutes with the target contrast, with stimuli varying along two relevant dimensions (coda closure duration, and vowel duration) to go from 'rose' to 'roads'. The cue-weighting task administered before and after training revealed mislabeling issues that were present *before* training: The language learners associated a short vowel with the word 'rose' instead of the word 'roads'. However, the learners improved their use of both vowel duration and coda closure duration towards native-like performance post training. Hence, discrimination training was effective for altering Japanese speakers' use of the acoustic cues that contrast the English words 'rose' and 'roads'.

INTRODUCTION

Living in an English-speaking country for an extensive period of time is no guarantee that adult learners of English will learn to distinguish novel speech sounds (e.g., Oh et al., 2011). Conversely, few hours of phonetic training appear to lead to the creation of new speech categories for a considerable number of adult learners (Grenon, Kubota, & Sheppard, 2019). Most phonetic training programs feature an identification task consisting of presenting second language (L2) learners with one word (aurally), for instance *ship*, and asking them to identify which word they heard, 'ship' or 'sheep'. This type of training is designed to improve identification accuracy and has been shown to be effective in the training of a variety of L2 segmental (e.g., Logan, Lively, & Pisoni, 1991; Iverson, Hazan, & Bannister, 2005; Wang & Munro, 2004) and suprasegmental contrasts (e.g., Wang, Spence, Jongman, & Sereno, 1999).

However, it also presupposes that the learners are familiar with the L2 grapheme-phoneme correspondence. But what if they are not? The use of a discrimination task may serve as an alternative, for instance, for training beginner English learners of Russian or young Japanese learners of English with difficult L2 contrasts. In the AX discrimination task, the learner hears two words, for instance *ship* and *sheep*, and has to decide if the two words are the 'same' or 'different'. Hence, this task does not require familiarity with the L2 orthographic system.

Comparisons between the use of an identification task and a discrimination task have yielded contradictory results. While some studies conclude that both types of tasks are equally effective (Flege, 1995; Wayland & Li, 2008), others suggest that the identification task is superior to the discrimination task (Carlet & Cebrian, 2015; Cebrian, Carlet, Gavaldà, & Gorba, 2018; Nozawa, 2015; Shinohara & Iverson, 2018). For instance, Spanish/Catalan speakers learning English as an L2 were trained to perceive English vowels using nonsense words produced by multiple talkers combined with either an identification task, AX discrimination task, or transcription task (control group) (Carlet & Cebrian, 2015). Training consisted of five 30-min sessions. Before and after training, learners' ability to identify English vowels was evaluated with the use of an identification task. The improvement in identification accuracy on both trained and new minimal-pairs was found to be higher for the identification training than the discrimination training group.

In a previous study we found that one hour of discrimination training with the English vowels in 'ship' and 'sheep' was sufficient for improving Japanese speakers' reliance on spectral information, as assessed through a cue-weighting task (Grenon, Sheppard, & Archibald, 2018). The cue-weighting task also revealed that twenty-five percent of the learners exhibited mislabeling issues, where they associated the vowel /i/ with the word 'ship' instead of 'sheep'. Considering that in most of the previous studies cited above, an identification task (using different minimal pairs) was used to assess improvement rather than a cue-weighting task (using one minimal pair manipulated along different acoustic dimensions), it is possible that mislabeling issues have contributed to lower the observed effect of the discrimination training paradigm. In turn, these results have led to the conclusion that discrimination training was not as efficient as identification training for improving the perception of new sound contrasts.

While we have shown that mislabeling issues may need to be addressed when training with a vowel contrast using a discrimination task (Grenon, Sheppard, & Archibald, 2018), the current study investigated the effect of the same task when training with a new consonantal contrast. The same twenty Japanese participants who trained with the vowel contrast reported in the study above were trained with the coda consonant contrast as in 'rose' and 'roads' using an AX discrimination task. This contrast is of particular interest in providing information about the learning of new speech sounds, as the words 'rose' and 'roads' are generally thought to be homophonous by Japanese speakers because the fricative /z/ and the affricate /d^{z}/ are phones in free variation in Japanese (that is, unlike the vowel contrast, Japanese speakers are unaware that these words are contrastive in English). Hence, the current study was expected to shed further light on the underlying learning mechanisms involved when training with a discrimination task.

Research questions and hypotheses

The specific research questions addressed by this paper are: (1) do Japanese speakers encounter mislabeling issues when training with a discrimination task with the 'rose' and 'roads' contrast, and (2) do they improve their sensitivity to the critical acoustic cues that serve to distinguish those sounds. The results of the L2 trainees are compared with that of native English speakers to evaluate if any change in perception is moving towards native performance.

Based on previous research (e.g., Grenon, 2011), it was expected that before training the Japanese speakers' sensitivity to vowel duration to distinguish short and long vowels in their L1 would be used to contrast the 'rose' and 'roads' stimuli in the cue-weighting task, while they would not rely significantly on the closure duration. After training, it was expected that the Japanese speakers would be relying less on vowel duration while relying more on the closure duration.

METHOD

Participants

The participants were the same twenty native Japanese speakers who took part in the study by Grenon, Sheppard and Archibald (2018) (the results of an additional participant were excluded from analysis for intensive exposure to English during early childhood). They were all students at the University of Tokyo in Japan aged between 18 and 27 years old (M = 20) who had never stayed in an English-speaking country for more than 8 weeks (M = 1.7 week). They received a monetary compensation for their participation.

Fifty-four native English speakers from North America also participated in this experiment. The data of fourteen participants were discarded either because the participant had been exposed regularly to another language during early childhood, or he or she reported a history of speech or hearing impairment. The resulting forty participants were all students at the University of Victoria in Canada aged between 17 and 28 years old (M = 21). They received course credits for their participation.

All participants signed a consent form prior to their participation. The Japanese participants completed the pre-test, two discrimination training sessions and post-test all on different days over a 2 to 3-week period (note that half of the participants were trained on the 'ship' and 'sheep' vowel contrast discussed previously before training on the 'rose' and 'roads' contrast). The time elapsed between the last 'rose' and 'roads' training session and post-test ranged between 1 and 11 days (M = 5.05, St. dev. = 3.14). The English participants completed the pre-test only (the pre-test and post-test were identical).

Stimuli

Six 'rose' and six 'roads' samples produced by a female university student from the United-States were recorded with a Sony microphone (ECM-MS957) at 44,100Hz in a sound attenuated booth at the University of Tokyo directly to computer using Praat (Boersma & Weenink, 2017). The intensity of the initial recording was scaled to 70dB. The vowel duration in the six recorded 'rose' samples varied from 280ms to 306ms (M = 291ms). The vowel duration in the six recorded 'roads' samples were consistently shorter, and varied from 220ms to 267ms (M = 243ms), while the closure duration varied from 64ms to 88ms (M = 78ms). Although the /d/ may disappear in fluent speech in some English dialects (Roca & Johnson, 1999), it was present in all the recorded 'roads' samples.

From the six recorded 'rose', a clear exemplar was chosen (i.e., without any glottalization or other features that may interfere with the manipulations). Then, 40ms of closure duration was extracted from a 'roads' sample and inserted between the vowel and the word-final fricative in the word 'rose'. The closure duration was then modified from 0ms to 60ms in 7 steps of 10ms using a script for making a duration continuum (Winn, 2014). Using the same script, the vowel duration of each of the seven tokens was modified to vary from 210ms to 300ms in 4 equal steps of 30ms. The 28 resulting tokens are schematized in Figure 1, with spectrograms of token 1 and 28 presented in Figure 2.

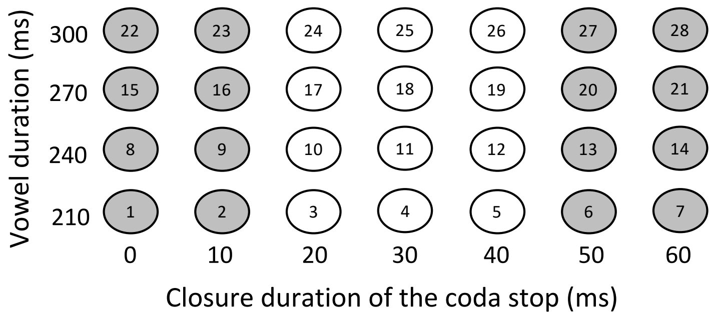


Figure 1. The 28 manipulated tokens used for the pre- and post-test were varied in terms of duration of the coda stop closure (x-axis) and vowel duration (y-axis). The 16 tokens used for training are presented in grey shading.

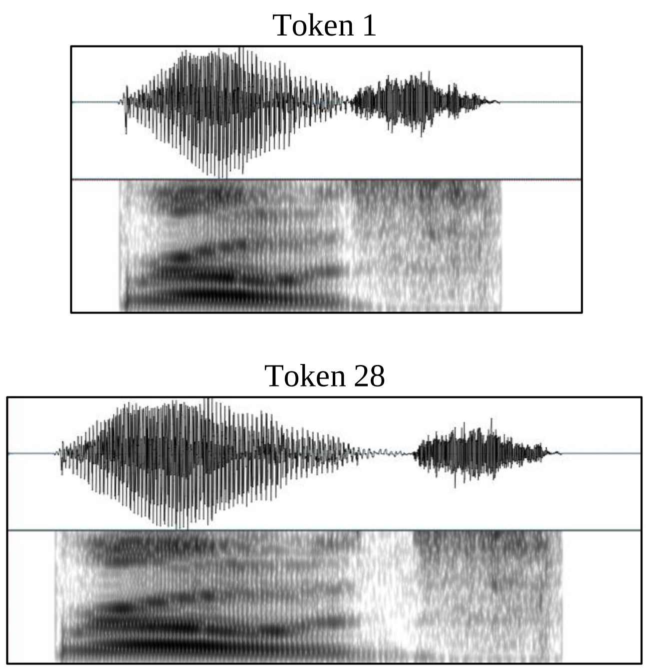


Figure 2. Spectrogram of the manipulated stimulus with no stop closure and a vowel duration of 210ms (token 1 in Figure 1). Spectrogram of the manipulated stimulus with 60ms of stop closure and a vowel duration of 300ms (token 28 in Figure 1).

The 28 resulting words were used in the identical pre- and post-test. A subset of 16 tokens were used for training. The tokens chosen for training were situated at the extreme ends of the closure duration continuum and are identified with grey shading in Figure 1. The 16 tokens were paired for the AX discrimination training task so that 16 combinations featured words that differed in terms of closure duration, such as token 2 in Figure 1 followed by token 6 (these should be labeled as 'different' by the participants), and 16 pairs featured words that may have different vowel duration, but the closure duration was within the same category, such as token 1 and token 16 (these should be labeled as 'same' by the participants). None of the words was paired with itself. The resulting 32 pairs were also presented in reverse order, for a total of 64 training pairs, presented randomly 4 times, for a total of 512 words heard during one training session.

Procedure

The pre- and post-test done in a sound-attenuated room were meant to evaluate the weighting of each acoustic cue manipulated. Before reading the set of instructions for the task in their respective L1, participants were required to wear BOSE AE2 headphones and adjust the sound level to the most comfortable setting. For the task, the participants were presented with a red cross in the middle of the computer screen for 1000ms, then heard one of the 28 manipulated tokens, and had to decide if the word was 'rose' or 'roads' (the written words appeared on the screen) by pressing the appropriate key on the response pad. No feedback was provided during the tests, the words were never repeated, and the learners were requested to respond as quickly as possible. The 28 manipulated tokens were presented randomly four times during a test for a total of 112 test tokens (the first round of 28 words, considered a practice session, was discarded from the analyses). A test lasted 5 to 10min, with no break.

After the pre-test and before the post-test, the Japanese listeners completed 2 training sessions of about 30min on two different days in a sound-attenuated room, for a total of 1h of discrimination training. For the training, the learner would hear two words with an inter-stimulus-interval of 1500ms (e.g., *rose – roads*), and had to decide if the words were the 'same' or 'different' (only the words 'same' and 'different' were written on the screen) by pressing the appropriate key on the computer keyboard. Each trial was followed by feedback (a written message) indicating whether the choice was correct. The next trial was presented after an inter-trial-interval of 2000ms added after a participant's response.

RESULTS AND DISCUSSION

Training

To assess any improvement during training, the raw training scores (percentages of correct responses) were computed for each participant. The average score on the first training session was near chance level, that is 52.40% (St. dev. = 11.07), meaning the contrast of interest was indeed very difficult for them. The Japanese participants slightly improved their performance on the second training session, with an average score of 56.02% (St. dev. = 16.22). Although small, this improvement was significant (t (20) = 2.483, $p < .05$, $d = 0.53$).

Mislabeling

To evaluate mislabeling issues, the proportion of tokens identified as 'roads' by native Japanese speakers were compared with that of native English speakers on each dimension of interest: vowel duration and closure duration of the stop coda consonant.

Mislabeling issues were found to occur *before* training on the vowel duration dimension as shown in Figure 3 below. That is, while English speakers associated a shorter vowel with the word 'roads' and a relatively longer vowel with the word 'rose', the Japanese trainees did the opposite, associating a shorter vowel with the word 'rose' instead. Given that the Japanese language distinguishes between short and long vowels, it is not surprising that Japanese participants are sensitive to vowel duration and may use it for distinguishing foreign words.

In the current case, the word 'rose' is generally the one with a longer vowel than the word 'roads' in English, and the results show that the English speakers in the study are sensitive to the vowel duration when contrasting the two words. For Japanese speakers, it is possible that the use of two vowel letters in the word 'r**oa**ds' has led them to mistakenly presume that this word contains a long vowel. Hence, an influence of orthography may be possible in this case. That being said, there was a change in the Japanese speakers' use of vowel duration between pre-test and post-test, where they started to rely less on vowel duration post training, as reported in the next subsection.

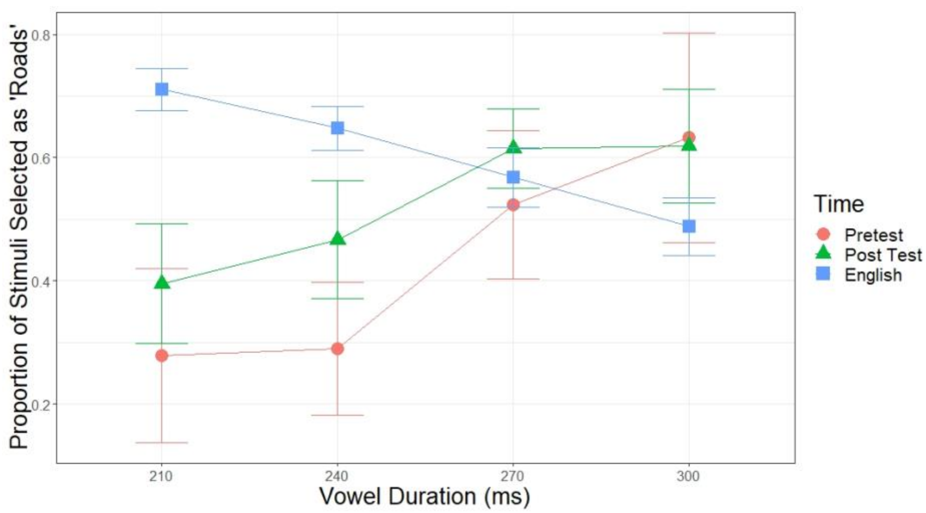


Figure 3. Results of the 28 test tokens across all vowel duration values for the Japanese listeners before (red circles) and after (green triangles) training compared with the results of the native English listeners (blue squares).

Cue-weighting

The results of a multi-level linear model analysis confirmed that the change in the use of vowel duration between pre-test and post-test was significant (X^2 (2) = 4.54, p < .05, d = 0.535). A follow up analysis comparing the Japanese native speakers' post-test performance with that of the native speakers found that the behavior of the Japanese speakers did not attain the same level as the native speakers as represented by a significant effect of Group (X^2 (2) = 6.65, p = .001, d = 0.714) as well as a significant Group x Vowel duration interaction (X^2 (4) = 92.11, p < .0001, effect sizes for the planned comparisons: d = 2.01-2.80).

After training, the L2 listeners also improved their sensitivity to the most critical acoustic cue that serves to distinguish the coda contrast. As shown in Figure 4, native English speakers rely heavily on the closure duration to categorize the word 'rose' and 'roads', whereas native Japanese listeners mostly ignored this cue before training. After training, however, their use of this cue has increased, and is starting to resemble the native speakers' performance.

The results of a multi-level linear model analysis confirmed that the identification behavior along the closure duration dimension of the L2 listeners changed significantly from pre-test to post-test (X^2 (2) = 4.94, p < .05, d = 0.521), although their performance was still different from native listeners with a significant effect of Group (X^2 (2) = 5.96, p < .05, d = 0.714) and Group x Closure duration interaction (X^2 (4) = 159.31, p < .0001, effect sizes of the planned comparisons: d = 1.35-2.57).

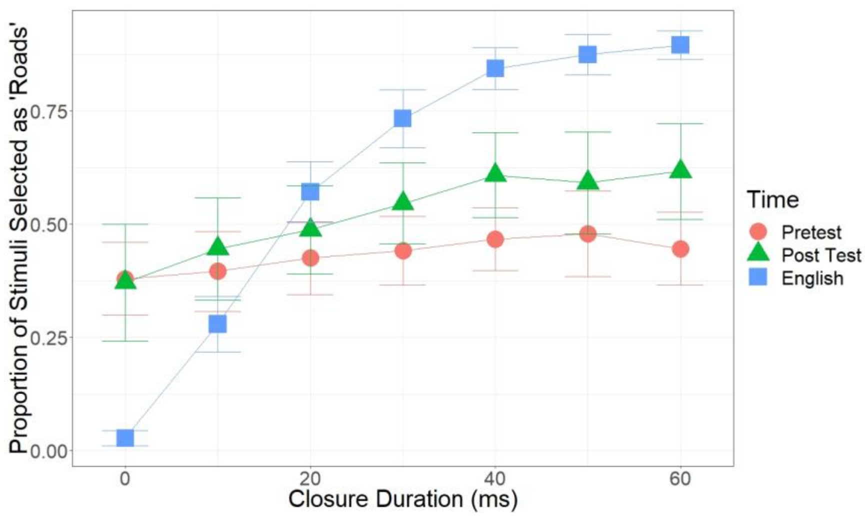


Figure 4. Results of the 28 test tokens across all closure duration values for the Japanese listeners before (red circles) and after (green triangles) training compared with the results of the native English listeners (blue squares).

Summary of results and general discussion

In sum, L2 listeners could change their cue-weighting for categorization of the 'rose' and 'roads' contrast towards native speakers' performance. After training, the Japanese listeners improved their use of both vowel duration and closure duration of the coda consonant for categorizing the coda contrast. In the case of the novel consonant contrast, mislabeling issues existed prior to the start of training, where listeners mistakenly associated a short vowel with the word 'rose' rather than with the word 'roads'. A possible effect of orthography was mentioned, where the use of one letter in the word 'rose' may have misled the L2 learners to assume that this word features a short vowel, whereas the word 'roads' features a long vowel. The possible effect of orthography could be ruled out by testing another minimal-pair featuring a single letter for the vowel sound in both words, such as 'cars' versus 'cards'.

A question that remains, however, is whether a change in the reliance of relevant acoustic cues would improve more with identification training. This possibility was tested using the same manipulated tokens presented the same number of times across two training conditions (identification vs. discrimination), with both the 'rose' and 'roads' contrast as well as with the 'ship' and 'sheep' contrast. The two tasks yielded comparable change in cue-weighting for the 'ship'-'sheep' contrast (Wee et al., in press) when disregarding mislabeling issues. Hence, the results of previous training studies with vowel sounds indicating that identification training is superior to discrimination training (Carlet & Cebrian, 2015; Cebrian et al., 2018; Nozawa, 2015) may have been affected by mislabeling issues.

However, for the 'rose'-'roads' contrast, identification training provided marginally superior results than discrimination training for the reliance on closure duration (Law et al., in press). Hence, while the two types of training were equally effective for the vowel contrast ('ship'-'sheep'), they were not as equally effective in the case of the consonantal contrast ('rose'-'roads'). A tentative explanation for this discrepancy, is that prior sensitivity to the critical acoustic cue may be required for discrimination training to be as effective as identification training. Also, in our studies, only one minimal-pair produced by one speaker was used. It remains to investigate if the use of high-variability combined with more training time could encourage further changes in cue-weighting.

CONCLUSION

The current study investigated whether a training program featuring a discrimination task could help native Japanese learners of English improve their sensitivity to the acoustic cues used by native English speakers to distinguish the coda consonants as in 'rose' and 'roads'. The results revealed that *before* training the Japanese trainees generally ignored the closure duration of the stop consonant, a cue that English speakers rely heavily on to contrast the word 'roads' from 'rose'. The L2 trainees relied instead on vowel duration to contrast the two words, but they mistakenly associated a short vowel with the word 'rose' instead of the word 'roads'. After training, however, their use of vowel duration developed towards native performance. Similarly, their use of the closure duration of the coda stop consonant improved after one hour of discrimination training, also developing towards native-like performance. Hence, discrimination training appears successful for improving the use of acoustic cues related to the perception of a consonant contrast when using a discrimination task.

ACKNOWLEDGMENTS

We are grateful to our many research assistants and participants without whom this research would not have been possible, and thankful also to Jim Tanaka. This research has received a grant-in-aid to scientific research by the Japan Society for the Promotion of Science, JSPS (16K02915) granted to Izabelle Grenon.

ABOUT THE AUTHORS

Izabelle Grenon is an associate professor in the Department of Language and Information Sciences at the University of Tokyo in Japan. Her research interests are in experimental phonetics related to first or second language acquisition. Her current research focuses on testing different phonetic training paradigms for improving the perception of non-native speech sounds.

E-mail: grenon@boz.c.u-tokyo.ac.jp

Chris Sheppard is an associate professor at the Center for English Language Education in Science and Engineering at Waseda University in Tokyo, Japan. His current research projects focus on fostering autonomy in English for Specific Purposes, the development of critical thinking through second language education, the acquisition of rhythm by learners of Japanese as a second language, and the role of feedback in second language acquisition.

E-mail: chris@waseda.jp

John Archibald is a professor of linguistics at the University of Victoria in Canada. He specializes in second language phonology. He has two current projects in the domain of intelligibility, one focusing on modified input, the other on psycholinguistic models of word recognition.

E-mail: johnarch@uvic.ca

REFERENCES

Boersma, P., & Weenink, D. (2017). Praat: doing phonetics by computer [Computer program]. Version 6.0.29, retrieved May 2017 from http://www.praat.org/.

Carlet, A., & Cebrian, J. (2015). Identification vs. discrimination training: Learning effects for trained and untrained sounds. *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS)*, Glasgow, Scotland, Aug. 10-14.

Cebrian, J., Carlet, A., Gavaldà, N., & Gorba, C. (2018). Effects of perceptual training on vowel perception and production and implications for L2 pronunciation teaching. Paper presented at *Pronunciation in Second Language Learning and Teaching (PSLLT)*, Ames, Iowa, Sept. 6-8.

Flege, J. E. (1995). Two procedures for training a novel second language phonetic contrast. *Applied Psycholinguistics*, *16*, 425-442.

Grenon, I. (2011). Dissociable levels of speech processing in second language perception. *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS)*, Hong Kong, Aug. 17-21.

Grenon, I., Kubota, M., & Sheppard, C. (2019). The creation of a new vowel category by adult learners after adaptive phonetic training. *Journal of Phonetics*, *72*, 17-34.

Grenon, I., Sheppard, C., & Archibald, J. (2018). Discrimination training for learning sound contrasts. Proceedings of *The 2nd International Symposium on Applied Phonetics (ISAPh)*, 51-56.

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. *Journal of the Acoustical Society of America*, *118* (5), 3267-3278.

Law, I. L. G., Grenon, I., Sheppard, C., & Archibald, J. (in press). Which is better: Identification or discrimination training for the acquisition of an English coda contrast. *Proceedings of the 19th International Congress of Phonetic Sciences (ICPhS).* Melbourne, Australia.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America*, *89* (2), 874-886.

Nozawa, T. (2015). Effects of attention and training method on the identification of American English vowels and coda nasals by native Japanese listeners. *Proceedings of the 18th International Congress of Phonetic Science (ICPhS)*, Glasgow, Scotland, Aug. 10-14.

Oh, G. E., Guion-Anderson, S., Aoyama, K., Flege, J. E., Akahane-Yamada, R., & Yamada, T. (2011). A one-year longitudinal study of English and Japanese vowel production by Japanese adults and children in an English-speaking setting. *Journal of Phonetics*, *39*, 156-167.

Roca, I., & Johnson, W. (1999). *A Course in Phonology*. Malden, MA: Blackwell Publishers.

Shinohara, Y., & Iverson, P. (2018). High variability identification and discrimination training for Japanese speakers learning English /r/-/l/. *Journal of Phonetics*, *66*, 242-251.

Wang, X., & Munro, M. J. (2004). Computer-based training for learning English vowel contrasts. *System, 32*, 539-552.

Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America*, *106* (6), 3649-3658.

Wayland, R. P., & Li, B. (2008). Effects of two training procedures in cross-language perception of tones. *Journal of Phonetics, 36* (2), 250-267.

Wee, D. T. J., Grenon, I., Sheppard, C., & Archibald, J. (in press). Identification and discrimination training yield comparable results for contrasting vowels. *Proceedings of the 19th International Congress of Phonetic Sciences (ICPhS).* Melbourne, Australia.

Winn, M. (2014). Make duration continuum [Praat script]. Version August 2014, retrieved April 14, 2017 from http://www.mattwinn.com/praat.html.

Henrichsen, L. (2019). An analytical instrument for evaluating computer-assisted pronunciation teaching software, websites, and mobile apps. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Ames, IA, September 2018 (pp. 137-146). Ames, IA: Iowa State University.

PRESENTATION/POSTER

AN ANALYTICAL INSTRUMENT FOR EVALUATING COMPUTER-ASSISTED PRONUNCIATION TEACHING SOFTWARE, WEBSITES, AND MOBILE APPS

Lynn Henrichsen, Brigham Young University

Computer-assisted pronunciation teaching (CAPT) offers many potential benefits—a private, stress-free learning environment; virtually unlimited input; practice at the student's own pace; individualized, instantaneous feedback through Automatic Speech Recognition (ASR); and visual acoustic and articulatory displays. Many CAPT programs, websites, and mobile apps have been created in recent years. Regrettably, however, CAPT software does not always measure up to its potential. Furthermore, many L2 teachers and learners, not familiar with the full range of CAPT possibilities, may not be aware of what features to look for in a CAPT product. This paper shares a remedy to this problem—a comprehensive set of criteria for analyzing and evaluating CAPT software, websites, and mobile apps. Utilizing an easy-to-use checklist, as well as Likert-scale and open-response items, of features that potential CAPT users should look for in software, this two-page instrument guides teachers and learners evaluating CAPT programs to consider variables and criteria recommended by pronunciation and CALL experts.

THE PROMISES AND LIMITATIONS OF CAPT

Pronunciation teachers and learners have long been enticed by the potential of computer-assisted pronunciation teaching (CAPT). Over 15 years ago, Neri, Cucchiarini, Strik, and Boves (2002, p. 441) stated that CAPT "can be beneficial to second language learning as it provides a private, stress-free environment in which students can access virtually unlimited input, practice at their own pace and, through the integration of Automatic Speech Recognition (ASR), receive individualized, instantaneous feedback."

Levis (2007) later made a similar point:

> The use of computers is almost ideally suited to learning pronunciation skills. Computers can provide individualized instruction, frequent practice through listening discrimination and focused repetition exercises, and automatic visual support that demonstrates to learners how closely their own pronunciation approximates model utterances. (p. 184)

More recently, others (Chun, 2013; Fouz-González, 2015) have described promising CAPT tools, such as visual acoustic displays (i.e., waveforms, spectrograms, pitch contours, and formant data), visual articulatory displays (i.e., sagittal section diagrams and still and video images of a speaker's mouth and lip movements), and automatic speech recognition (ASR).

Today, increasing numbers of language learners and teachers are coming to rely on websites and mobile apps to help them improve their English language skills (Rosell-Aguilar, 2017, p. 243). A large number of online programs have been developed with the goal of helping English language learners (and their teachers) with various aspects of English—including pronunciation.

Regrettably, not all these websites and apps are as helpful or effective as they could be. For this reason,

> Teachers and learners should not be seduced by the strong appeal of the marketing done by publishers. Instead, it is necessary to analyze English as a Foreign Language and/or Second Language (EFL/ESL) pronunciation teaching software programs as to their potential for developing English pronunciation. There is an unquestionable need to analyze these programs from a critical perspective using pedagogically coherent and technically elaborated criteria. (Martins, Levis, & Borges, 2016, p. 142)

O'Brien and Levis echo this warning: "Many of the commercially available products are often neither pedagogically sound nor informed by research" (2017, p. 1). Neri, et al. (2002, p. 441) also laud the "wealth of CAPT systems" available but caution,

> When examined carefully…the display of products may not look entirely satisfactory. Many authors describe commercially available programs as fancy-looking systems that may at first impress student and teacher alike, but eventually fail to meet sound pedagogical requirements.…These systems, which do not fully exploit the potentialities of CAPT, look more like the result of a technology push, rather than of a demand pull. (p. 442)

More recently, Kaiser (2017, slide 45) has noted that instruction is often based more on "what is easier to program or what will 'sell' the app than what is best pedagogy." Rather than giving primacy to proven pronunciation-teaching/learning principles and procedures, "many apps have been developed with more attention to appearance and flash" (Yoshida, 2018, p. 208).

In sum, while initially offering marvelous promise, CAPT software has not always delivered on that promise. Instead, it has frequently failed to embody the ideals envisioned for it.

CHARACTERISTICS OF IDEAL CAPT PROGRAMS

Many experts have made recommendations for the design of high quality CAPT software. For instance, considering not only the pedagogical but also the technological aspects of CAPT software, Neri, et al. formulated a set of recommendations for model CAPT programs:

> Learning must take place in a stress-free environment in which students can be exposed to considerable and meaningful input, are stimulated to actively practice oral skills and can receive immediate feedback on individual errors. Input should pertain to real-world language situations, it should include multiple-speaker models and it should allow the learner to get a sense of the articulatory movements involved in the production of L2 speech. Oral production should be elicited with realistic material and exercises catering for different learning styles, and should include pronunciation of full sentences. Pertinent and comprehensible feedback should be provided individually and with minimum delay and

> should focus on those segmental and suprasegmental aspects that affect intelligibility most. (2002, p. 449)

Today, CAPT users are still hoping for programs that have all (or even most) of the above features—especially automatic speech recognition (ASR) that will "recognize everything the user says, point out those areas that are most problematic (depending on the user's priorities, be it intelligibility, comprehensibility or accuracy), and then offer explicit feedback indicating how to improve" (Fouz-González, 2015, p. 324). Such instantaneous, individualized, targeted feedback is desirable as it will lead to greater learner autonomy, responsibility, and self-monitoring on the part of L2 pronunciation learners. Discussing the use of ASR in CAPT, McCrocklin (2016) made the case that students should become autonomous learners of pronunciation by developing skills and using strategies that will enable them to practice pronunciation on their own and not rely on a teacher for pronunciation training. For this reason, she praised online resources as tools that can potentially "promote autonomy by enabling experimentation through self-access work outside of class while also providing immediate feedback to learners" (p. 27).

Putting all these desirable criteria together forms an impressive CAPT-software wish list: A stress-free learning environment, meaningful and realistic input, active practice beyond the word level, instant and individualized corrective feedback that is targeted on specific perception or production problems, multiple speaker models, articulatory explanations and visualizations, allowances for different learning styles and rates, focus on the aspects of pronunciation—both segmental and suprasegmental—that are most important to intelligibility, and the promotion of greater autonomy and more effective strategy use on the part of L2 pronunciation learners.

CURRENT CAPT REALITY

The ideal CAPT characteristics listed above are clearly desirable, but in the real world today pronunciation-focused websites and mobile apps rarely measure up to all these criteria and expectations. Often a significant gap exists between current, research-based pronunciation theory and pedagogy and actual CAPT applications. Lamentably, many appear "suspiciously like traditional, drill-oriented pedagogy in new clothing" (Levis, 2007, p. 185). Kaiser (2017) recently analyzed 30 L2 pronunciation teaching/learning apps and found that, of the 30 apps, 22 (73.3%) relied heavily on a simple listen-and-repeat instructional approach and provided no feedback to learners regarding the accuracy of their production. Only a few apps provided visual feedback in the form of spectrograms. Regarding the eagerly anticipated, long-promised benefits of ASR, Kaiser's (2017) analysis determined that the few mobile apps he examined that employed automatic-speech-recognition provided simplistic, dichotomous "correct" or "incorrect" feedback that was not necessarily accurate. From this and other sources of evidence, it appears that in spite of the great potential that ASR holds for providing automatic feedback on learners' pronunciation, it…

> Needs to improve substantially before learners can use these systems autonomously and rely entirely on their judgments. The effectiveness of these systems decreases significantly when dealing with non-native speech…and ASR ratings do not always correlate with those by human raters.…In spite of advances in the field, an acceptable level of reliability is only guaranteed when the tasks are simple and utterances are kept to a restricted set from which

> students select a response…something that limits the usability of this technology for spontaneous practice. (Fouz-González, 2015, p. 328)

Fouz-González (2015) also points out that when used with L2-accented speech, ASR typically produces "numerous false alarms and low rates of correct detection," resulting in an experience that "may be quite frustrating for users if mistakes are not detected or are detected incorrectly." Further, "once learners suspect the system is not reliable, they will lose confidence in it" (p. 328). Thus, using inadequate ASR software can "lead to frustrating and counter-productive experiences if learners waste time trying to match a model when their pronunciation is already acceptable" (p. 327). In sum, ASR software has still not reached the point where it provides reliable feedback to L2 learners. On the bright side, in recent years, the speech-to-text application programming interfaces (APIs) —such as *IBM Watson*, *Google Voice*, and *CMU Sphinx*—underlying commercial voice-recognition software and AI interfaces—such as *Siri, Alexa, Google Home*, and *Amazon Echo*—have become increasingly powerful and accurate. Consequently, the accuracy of ASR dictation systems has increased for both native and non-native speakers (McCrocklin, Humaidan, & Edalatishams, 2018). In the future, the evaluation validity and reliability of ASR in pronunciation apps built with these improved APIs will surely do likewise.

For all these pedagogical and technological reasons, L2 teachers and learners must exercise caution when selecting CAPT software, and they must consider a variety of criteria. Some programs may have initial appeal because of an attractive feature but be lacking in other, important ways. For instance, some L2 pronunciation websites and apps may provide articulatory explanations but no practice. Others might require an accompanying teacher or textbook since they provide practice but minimal explanation or guidance for learners.

Many other differences exist among online resources for L2 pronunciation teaching and learning. For instance, some are free, while others require users to pay a fee. Some focus on segmentals, others on suprasegmentals, and some provide instruction and practice with both. In their user interface, some sites or apps provide helpful graphics, some contain only text, and a few provide video clips that help learners both see and hear how to pronounce English sounds correctly. Some programs follow a flexible, individualized approach, while others expect every user to follow the same curricular path. In sum, the number of criteria to consider when evaluating or selecting CAPT software is considerable, and potential users who focus on some features or criteria while overlooking others do so at their peril. Choosing faulty or inadequate software can lead to frustration and diminished learning experiences.

The analytical instrument described in the next section of this paper is intended to help language teachers and learners avoid such problems by providing a comprehensive set of criteria for analyzing CAPT software, websites, and mobile apps.

THE ANALYTICAL INSTRUMENT FOR EVALUATING CAPT SOFTWARE

Derwing and Munro (2015, p. 124) urge teachers evaluating, selecting, or recommending CAPT software to "read reviews and recommendations from authoritative sources and then to screen apps carefully before recommending them to students." The instrument described here can be used for conducting such screening. It includes information that will guide teachers and learners of L2 pronunciation in selecting the most appropriate and helpful CAPT resources for their learning/teaching needs. (It might even motivate CAPT software producers to create better products.) It provides a two-page listing of characteristics that potential users of CAPT software, websites, or mobile apps should look for and evaluate before deciding on a particular instructional product.

This instrument has been developed over many years. It began with Persichitte's (1995) "Basic Criteria for Selecting and Evaluating Instructional Software" and was later expanded with elements from Epstein and Ormiston's "Criteria for Developing and Evaluating Materials" (2007, pp. 9-10). Over time, pronunciation-specific criteria advocated by experts (Derwing & Rossiter, 2002; Martins, Levis, & Borges, 2016; Morley, 1991; Munro & Derwing, 2006; Neri, Cucchiarini, Strik, & Boves, 2002; Rosell-Aguilar, 2017) were added. The result was a two-page listing of characteristics or criteria that potential CAPT users should consider when evaluating a particular software product for pronunciation teaching or learning.

For many years, I have used the different, pilot versions of this instrument to evaluate language-teaching software. In the last few years, after refining and focusing the instrument, my graduate students and I have successfully used these criteria for evaluating CAPT software (Henrichsen, 2019; Henrichsen et al., 2018).

Figure 1 shows the front of the two-page evaluative instrument, and Figure 2 shows the reverse side.

Criteria for Analyzing and Evaluating Computer-Assisted Second-Language Pronunciation-Teaching and Learning Software, Websites, and Mobile Apps

A. General descriptive information

1. Name of evaluator: ________________________ 2. Date of evaluation: _______________
3. Software title: ____
4. Copyright (or last update) date of software: _____
5. Author, sponsor, or publisher's name (and qualifications): __________________________________
6. Platform: ❏ iOS, ❏ Android, ❏ Macintosh, ❏ Windows, ❏ Linux, ❏ other _______________
7. Target language(s): __________ (If English: ❏ British, ❏ North American, ❏ other ______)
8. Language(s) in which instructions are provided: ______________________________
9. Cost: $_______, ❏ free, ❏ ads, ❏ subscription, ❏ other ____________
10 Target audience age(s): ❏ children, ❏ teenagers, ❏ adults, ❏ other__________
11. Target audience language level(s): ❏ novice, ❏ intermediate, ❏ advanced, ❏ other ______

B. Instructional purpose(s) and activities

1. Primary objectives specified by program (e.g., focuses on developing intelligible [not native-speaker] pronunciation):____
2. Other (secondary, peripheral) objectives: ____
3. Aspects of pronunciation addressed (check all that apply):
   Listening perception: ❏ segmentals, ❏ suprasegmentals, other _______
   Segmentals: ❏ vowels, ❏ consonants, ❏ consonant clusters, other _________
   Suprasegmentals: ❏ intonation, ❏word stress, ❏ sentence stress, ❏ rhythm, ❏ pausing/juncture, ❏ blending, ❏ reduction, ❏ other ________
   Fluency: ❏ pausing appropriately, ❏ delivery speed, ❏ other _______
4. Type(s) of learning activities provided (check all that apply):
   ❏ Analysis or diagnosis of learner's difficulties
   ❏ Listening perception or discrimination
   ❏ Listen to a model and imitate/repeat
   ❏ Listen, record, replay, listen, and compare to model
   ❏ Minimal pairs
   ❏ Variable input (using speech models of different genders, regional dialects, registers, etc.)
   ❏ Contrasts between the learners' L1 sounds and corresponding L2 sounds
   ❏ Loaded sentences or "tongue twisters"
   ❏ Phonetic alphabet symbols and/or charts
   ❏ Articulatory explanations (text or video)
   ❏ Articulatory displays (sagittal section diagrams)
   ❏ Animated "mouth movement" models (video)
   ❏ Visual pitch-contour displays
   ❏ Spectrograms, waveforms, or formant data (of model and/or learner's pronunciation)
   ❏ Flash cards
   ❏ Instructional game
   ❏ Pair-work or group-collaboration activities
   ❏ Automatic speech recognition (ASR) How accurate is the ASR? How often (___%) does it reject correct/acceptable pronunciation as incorrect? How often (___%) does it accept incorrect pronunciation as correct? How does it do with different dialects?
5. Feedback and record keeping:
   ❏ Provides users with immediate feedback on the correctness of their responses
   ❏ Provides pronunciation feedback that is easily understandable by L2 learners
   ❏ Provides users with helpful, comprehensible guidance on how to correct their mistakes
   ❏ Tracks the number of right and wrong responses for an individual user
   ❏ Keeps track of various individuals in one class and reports scores to a teacher

Figure 1. First page of system for analyzing and evaluating CAPT software.

Rate each of the items in sections C, D, and E according to the following scale:

-2	-1	0	1	2
strongly disagree	disagree	no opinion	agree	strongly agree

NA=Does not apply CT=Cannot tell (insufficient data)

Write comments anywhere they fit or on a separate sheet (please refer to item numbers).

C. Functionality and usability

1. Runs properly (i.e., no bugs, crashes, long delays, etc.)	-2	-1	0	1	2	NA	CT
2. Guides the user well (i.e., intuitive interface, provides clear directions for starting, navigating, and stopping)	-2	-1	0	1	2	NA	CT
3. Uses consistent commands and directions throughout	-2	-1	0	1	2	NA	CT
4. Provides operational "Help" for users	-2	-1	0	1	2	NA	CT
5. Allows users to provide feedback or ask questions to the creators	-2	-1	0	1	2	NA	CT

D. Instructional factors

1. Presents information well (i.e., clearly, concisely, interestingly etc.)	-2	-1	0	1	2	NA	CT
2. Provides adequate, thorough, and effective practice	-2	-1	0	1	2	NA	CT
3. Provides helpful feedback	-2	-1	0	1	2	NA	CT
4. Focuses on priority aspects of pronunciation (e.g., functional load)	-2	-1	0	1	2	NA	CT
5. Content is authentic, up to date, and accurate	-2	-1	0	1	2	NA	CT
6. Presents speech in contexts (not just unrelated, individual words)	-2	-1	0	1	2	NA	CT
7. Provides various speech models (i.e., multiple speakers' voices)	-2	-1	0	1	2	NA	CT
8. Delivers instruction at a level appropriate for the target audience	-2	-1	0	1	2	NA	CT
9. Maintains a constant (or gradually increasing) level of difficulty	-2	-1	0	1	2	NA	CT
10. Presents teaching/learning activities in a good sequence	-2	-1	0	1	2	NA	CT
11. Provides helpful interaction with the user(s)	-2	-1	0	1	2	NA	CT
12. Allows for learner autonomy and independence	-2	-1	0	1	2	NA	CT
13. Allows users to repeat activities they have difficulty with	-2	-1	0	1	2	NA	CT
14. Allows individualization (learners choose which pronunciation features to work on)	-2	-1	0	1	2	NA	CT
15. Provides meaningful practice (using words learners know)	-2	-1	0	1	2	NA	CT
16. Provides communicative practice (bridging an information gap)	-2	-1	0	1	2	NA	CT
17. Provides variety in practice activities	-2	-1	0	1	2	NA	CT
18. Promotes metacognitive activity regarding pronunciation	-2	-1	0	1	2	NA	CT
19. Encourages learners to take responsibility for their improvement	-2	-1	0	1	2	NA	CT
20. Encourages learner strategy development	-2	-1	0	1	2	NA	CT
21. Supports a variety of learning styles (e.g., visual, auditory, etc.)	-2	-1	0	1	2	NA	CT

E. Presentation (User interface)

1. Uses appropriate, readable text (size, style, variety, and continuity)	-2	-1	0	1	2	NA	CT
2. Avoids distracting elements (unnecessary sounds, animations, ads)	-2	-1	0	1	2	NA	CT
3. Is not too busy or confusing (e.g., employs "white space" appropriately)	-2	-1	0	1	2	NA	CT
4. Utilizes an attractive, appropriate color scheme	-2	-1	0	1	2	NA	CT
5. Is aesthetically pleasing in general and looks professional	-2	-1	0	1	2	NA	CT
6. Audio clarity level is high	-2	-1	0	1	2	NA	CT
7. Audio volume is adequate and adjustable	-2	-1	0	1	2	NA	CT
8. Audio can be played at different speeds (fast, slow)	-2	-1	0	1	2	NA	CT

F. Summary

1. Strong points?
2. Weak points?
3. Other comments?

Figure 2. Second page of system for analyzing and evaluating CAPT software.

Page one consists of "fill in the blank" or "check" items focusing on (A) general descriptive information and (B) instructional purposes and activities. For example, the final item under item B-4 asks specific questions about automatic speech recognition (ASR), such as "How often (___%) does it reject correct/acceptable pronunciation as incorrect?"

On the second page, descriptive statements regarding (C) functionality and usability, (D) instructional factors, and (E) presentation are evaluated using a five-point scale with additional *Does not apply* and *Cannot tell* options. For instance, item D-7 "Provides various speech models [i.e., multiple speakers' voices]" refers to the benefits of high variability phonetic training (Bradlow, 2018; Wang & Munro, 2004). The wide-open items in the "Summary" section give evaluators total freedom in describing anything they see as the overall strengths and weaknesses of the software.

CONCLUSION

My TESOL MA students and I have found the guidance this instrument provides to be very beneficial as it focuses our attention on the many characteristics that need to be considered when evaluating CAPT products. By drawing our attention to factors that we might otherwise overlook, it results in a more thorough and professional analysis of the CAPT software under examination. I offer it here for the same reasons. I hope you, your students, and your colleagues will find it helpful when you need to evaluate CAPT websites or mobile apps.

ABOUT THE AUTHOR

Lynn E. Henrichsen (Ed.D., University of Hawai'i at Manoa) is a professor of TESOL in the Linguistics Department at Brigham Young University in Provo, Utah. He teaches courses in instructional methods (including the teaching of listening, speaking and pronunciation), research methods, phonetics/phonology, and ESL pronunciation. He is the author of *Pronunciation Matters* (University of Michigan Press, 1999, now available at www.pronunciationmatters.com). In his 45 years as a language teacher, he has taught Spanish and English language learners and teachers and conducted TESOL-related research in 10 different countries around the world. He has authored over a dozen books, 30 chapters in books, and 70 articles in professional journals.

4040 JFSB
Department of Linguistics
Brigham Young University
Provo, UT 84602
USA
801.422.2938
Lynn_Henrichsen@byu.edu

REFERENCES

Bradlow, A. (2018, September). High variability training in the lab and in the language classroom. Plenary presentation at the 10th annual PSLLT (Pronunciation in Second Language Learning and Teaching) Conference, Ames, IA.

Chun, D. M. (2013). Computer-assisted pronunciation teaching. In C. A. Chapelle (Ed.), *Encyclopedia of applied linguistics* (pp. 823-834). Oxford: Wiley-Blackwell. doi: 10.1002/9781405198431.wbeal0172

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

Derwing, T. M., & Rossiter, M. J. (2002). ESL learners' perceptions of their pronunciation needs and strategies. *System*, *30,* 15l5-166.

Epstein, R., & Ormiston, M. (2007). *Tools and tips for using ELT materials: A guide for teachers.* Ann Arbor: University of Michigan.

Fouz-González. J. (2015). Trends and directions in computer-assisted pronunciation training. In J. A. Mompean & J. Fouz-González (Eds.), *Investigating English pronunciation: Trends and directions* (pp. 314-342). Basingstoke & New York: Palgrave Macmillan.

Henrichsen, L. (2019, February). 21 online pronunciation resources for teaching and learning. *TESOL Connections*. Retrieved from http://newsmanager.commpartners.com/tesolc/issues/2019-02-01/2.html

Henrichsen, L., Devenport Blanco, K., Carreño, S., Carter, S., Decker, L., Fry, L.…Zhao, K. (2018). Online resources for learners and teachers of English language pronunciation. *TESL Reporter, 51*(1), 23-89. Retrieved from http://ojs-dev.byuh.edu/index.php/Issue1/article/view/1183/1134

Kaiser, D. J. (2017). iPronounce: Understanding pronunciation apps. Online Webinar, 8 June 2017. Recording and slides available at http://ollren.org/events/past-events#s-lg-box-wrapper-17792711

Levis, J. M. (2007). Computer technology in teaching and researching pronunciation. *Annual Review of Applied Linguistics*, 27, 184-202

Martins, C. G. de F. M., Levis, J. M., & Borges, V. M. C. (2016). The design of an instrument to evaluate software for EFL/ESL pronunciation teaching. *Ilha do Desterro, 69*(1), 141-160. doi: *http://dx.doi.org/10.5007/2175-8026.2016v69n1p141*

McCrocklin, S. (2016). Pronunciation learner autonomy: The potential of automatic speech recognition. *System*, *57,* 25-42. http://dx.doi.org/10.1016/j.system.2015.12.013

McCrocklin, S., Humaidan, A., & Edalatishams, I. (2018, September). ASR dictation programs accuracy: Have current programs improved? Paper presented at the 10th annual PSLLT (Pronunciation in Second Language Learning and Teaching) Conference, Ames, IA.

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

Munro, M. J., & Derwing, T. M. (2006). The functional load principle in ESL pronunciation instruction: An exploratory study. *System, 34,* 520-531.

Neri, A., Cucchiarini, C., Strik, H., & Boves, L. (2002). The pedagogy-technology interface in computer assisted pronunciation training. *Computer Assisted Language Learning, 15*(5), 441-467.

O'Brien, M. G., & Levis, J. M. (2017). Pronunciation and technology. In M. O'Brien & J. Levis (Eds.), *Proceedings of the 8th Pronunciation in Second Language Learning and Teaching Conference* (pp. 1-9). Ames, IA: Iowa State University. https://apling.engl.iastate.edu/alt-content/uploads/2017/05/PSLLT_2016_Proceedings_finalB.pdf

Persichitte, K. (1995). Basic criteria for selecting and evaluating instructional software. In D. A. Willis, B. Robin, & J. Willis (Eds.), *Technology and teacher education annual, 1995: Proceedings of SITE 95—Sixth International Conference of the Society for Information Technology and Teacher Education (SITE)* (pp. 379-382). Charlottesville, VA: Association for the Advancement of Computing in Education.

Rosell-Aguilar, F. (2017). State of the app: A taxonomy and framework for evaluating language learning mobile applications. *CALICO Journal, 34.2*, 243-258. doi: https://doi.org/10.1558/cj.27623

Wang, X., & Munro, M. J. (2004). Computer-based training for learning English vowel contrasts. *System, 32*, 539-552. doi: 10.1016/j.system.2004.09.011

Yoshida, M. T. (2018). Choosing technology tools to meet pronunciation teaching and learning goals. *CATESOL Journal, 30*(1), 195-212.

Inceoglu, S. (2019). Exploring the perception and production of L2 French vowels: The role of phonological memory. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Ames, IA, September 2018 (pp. 147-157). Ames, IA: Iowa State University.

PRESENTATION/POSTER

EXPLORING THE PERCEPTION AND PRODUCTION OF L2 FRENCH VOWELS: THE ROLE OF PHONOLOGICAL MEMORY

Solène Inceoglu, Australian National University

This study explored how second language (L2) speech production of French nasal vowels is related to both L2 speech perception and individual cognitive differences in phonological short-term memory (PSTM) and working memory (WM). Thirty-two Australian-English native speakers enrolled in French language courses completed a delayed-repetition task to assess their production and a set of force-choice identification tasks in audiovisual (AV), audio-only (A) and visual-only (V) modalities to measure their perception skills. They then completed a non-word repetition task assessing their PSTM, and a listening span test measuring their WM. Results revealed that accurate production scores were higher for [ɛ̃] (91%), followed by [ɑ̃] (60%), and [ɔ̃] (55%), and that the perception and production scores were strongly correlated (AV: r = .66, A: r = .65, V: r = .68, all with large effect sizes). In terms of individual differences, there was a significant effect of PSTM on production and perception scores, but no effect of working memory capacity. The results are discussed in relation to current research on PSTM and L2 phonology, and with reference to theoretical and pedagogical implications.

INTRODUCTION

Some second language (L2) learners are more successful than others at perceiving and producing L2 contrasts. Variability across individuals can be attributed to a number of factors such as language proficiency, age of L2 learning, and type of instruction, among others. However, variation in the degree of success in L2 learning can also be further explained by differences in general cognitive abilities, including working memory (WM), phonological short-term memory (PSTM), and processing speed. To date, most SLA studies investigating cognitive individual differences have focused on the development of L2 grammar, vocabulary, and fluency (Kormos & Sáfár, 2008; Speciale, Ellis, & Bywater, 2004; Williams & Lovatt, 2003), and little is known about their effects on L2 phonology. Accordingly, the goal of the current study was to further explore the contribution of WM and PSTM on the perception and production of L2 French nasal vowels.

Phonological short-term memory and working memory

One of the most influential conceptualization of WM today is the model developed by Baddeley and Hitch (1974). The model contains three main components each with limited capacities in terms of storage and processing and with relative interdependence: the central executive, acting as the general attentional controller and allocating the finite resources of the WM system, and two slave systems called the phonological loop and the visuospatial sketchpad. The visuospatial sketchpad handles the processing and storage of visual information, whereas the phonological loop deals with verbal and acoustic information and consists of a storage component that holds speech-based information for a few seconds unless the decaying information is refreshed by an articulatory rehearsal process. In a more recent model of WM, Baddeley (2000) added a fourth element—the episodic buffer—which serves as an interface between the short-term storage systems and long-term memory.

While both the executive component of WM and the phonological loop have been implicated in the development of second language acquisition, they are two different concepts and are measured with different tasks. WM, which emphasizes the processing, manipulation, and storage of information, is often assessed with complex tasks such as reading, listening, or speaking span tasks, and has been found to play an essential role in the domains of lexical development (Kroll, Michael, Tokowicz, & Dufour, 2002), grammatical processing (e.g., Coughlin & Tremblay, 2013), and reading skills (Harrington & Sawyer, 1992). High WM is also associated with L2 phonological development (Darcy & Mora, 2015) and enhanced L2 oral performance in complexity, accuracy, and fluency (Mota, 2003), but was not found to be correlated with pronunciation ratings of L2 Spanish (Posedel, Emery, Souza, & Fountain, 2012).

PSTM, on the other hand, is measured with non-word repetition or recognition tasks, digit span tasks, or serial recall tasks, and has been found to be implicated in the development of L2 vocabulary (e.g., Speciale et al., 2004) and grammar and morphosyntax (e.g., O'Brien, Segalowitz, Collentine, & Freed, 2006; Serafini & Sanz, 2016). Phonological memory also accounted for some of the variance associated with oral fluency development (O'Brien, Segalowitz, Freed, & Collentine, 2007) and oral production skills (French, 2006; Kormos & Sáfár, 2008), but very little work has been done specifically on the relationship between PSTM and L2 pronunciation accuracy. Notable exceptions are a study with L1 Japanese speakers (Kondo, 2012) reporting a positive influence of verbal and non-verbal phonological memory on L2 English pronunciation skills, and a study on individual differences in L1 English-L2 Spanish (Nagle, 2013) showing that PSTM correlated (r = .51, p = .002) with mean pronunciation rating. There seems to be, however, a growing interest in this domain as illustrated by recent conference presentations. For instance, Zahler and Lord (2018) found that high PSTM learners demonstrated acoustic properties that were more closely similar to those of native speakers, with less centralizing of unstressed vowels than low PSTM learners. PSTM abilities did not, however, affect vowel duration, with high and low PSTM learners producing much longer vowels than native speakers. In a similar vein, Kondo (2018) investigated the link between the L2 pronunciation skills of 70 Japanese learners of English and acoustic short-term memory measured with a Tonal Memory Span Test and a Rhythm Memory Span Test. Her results revealed that acoustic short-term memory had significant positive effects on English word reproduction skills, with stronger effects observed for tonal memory capacity.

Connections have also been found between PSTM and L2 speech perception. MacKay, Meador, and Flege (2001) investigated the relationship between PSTM and the perception of L2 English consonants by Italian native speakers and found that phonological memory accounted for 8% to 15% of the variance in identification scores. The advantage of higher PSTM was further demonstrated in a study of English consonant perception by L1 speakers of Greek (Lengeris & Nicolaidis, 2014) and a perceptual training study of L2 English vowels by Catalan/Spanish speakers (Aliaga-García, Mora, & Cerviño-Povedano, 2011). Research on the contribution of PSTM on L2 speech perception is, however, still scant and sometimes contradictory. For instance, PSTM was found to have an important influence on the perception of L2 English /iː/-/ɪ/ cue weighing by Catalan/Spanish bilinguals in Cerviño-Povedano and Mora's (2010) study, but this influence was not replicated in Safronova and Mora's (2012) study.

Research questions

Given the scarcity of research and often conflicting results on the relationship of WM and PSTM to L2 production and perception, this study aims to answer the following research questions:

- How do Australian English learners of French perceive and produce French nasal vowels?
- How is L2 production and perception of French nasal vowels related to individual cognitive differences in PSTM and WM?

METHODS

Participants

The participants were 32 Australian-English learners of French (4 male) between the age of 18 and 33 (mean age = 20.3). They were enrolled in several levels of French at a large Australian university and all had completed at least one full semester of French. None of the participants reported hearing or vision problems.

Production task and production rating

A delayed repetition task was used to collect participants' productions of the three French nasal vowels [ɔ̃-ɑ̃-ɛ̃]. Participants were presented with a total of 108 CVC stimuli where V was one of the nasal vowels followed by a prompt in French inviting them to repeat the word (e.g., "[pɑ̃d] *répète le mot s'il te plait*"). The rating of the participants' productions was conducted with a forced-choice identification task (Inceoglu, 2015) whereby the researcher, also a native speaker of French, listened to participants' recordings and chose which of the nasal vowels had been produced. A second native speaker of French rated 31% of the data (10 participants) with an interrater reliability of 97%. For more details regarding the stimuli, production task, and rating task, see Inceoglu (2016).

Perception tasks

The stimuli for the perception task were the same as those used by Inceoglu (2016). A total of 108 CVC word containing one of the three French nasal vowels [ɔ̃-ɑ̃-ɛ̃] were recorded by a female native speaker of French. The initial consonant was one of the following: [p-t-k-b-d-g-s-z-f-v-ʒ-ʃ] to take into consideration the articulation of vowels in different consonantal contexts. The perception task was administered in three modalities of presentation: audiovisual (AV), audio-only (A), and visual-only (V). The stimuli were the same for the three tasks but were randomized across tasks and participants. Participants heard the stimulus and were asked to identify the nasal vowel by clicking on one of the three options on the screen, <on> on the left, <an> in the middle, and <un> on the right (i.e., [ɔ̃-ɑ̃-ɛ̃], respectively). They had four seconds to make their selection before presentation of the next stimulus, and no feedback was provided. The experiment was conducted using the software program SuperLab (Cedrus, 2015) and was preceded by a practice task to familiarize participants with the procedure.

Phonological short-term memory task

Participants completed a non-word repetition test to assess their PSTM (e.g., Grey, Williams, & Rebuschat, 2015; Kissling, 2014; Lado, 2008). The test consisted of 16 pairs of English non-words, spoken by a female Australian-English speaker, with syllable lengths increasing from 3 (e.g., *melistok, nutolon*) to 8 (e.g., *towarymanitacorous, finterprofensibolities*). Directions were given aurally and in writing, and three additional pairs of non-words served as practice. Participants heard the 16 pairs and were asked to repeat each pair after a two-second delay tone. As the number of syllables increased, the task became more challenging. Participants were awarded one point for each pair that they repeated correctly with no more than one erroneous syllable, resulting in a maximum total of 32 points.

Working-memory task

Participants' WM capacity was evaluated with a sentence span test (Daneman & Carpenter, 1980; Winke, 2013). Participants were aurally presented with 48 unrelated sentences at 3-second intervals in sets of three, four, and five sentences. Half of the sentences were grammatically correct and half were semantically plausible, resulting in four sentence types. For each set, participants were asked to judge the grammaticality and plausibility of the sentences and recall the last word of each sentence before moving on to a new set. Half a point was awarded for a correct judgment of plausibility, half a point for a correct judgment of grammaticality, and one point for each correctly recalled word. This totaled a maximum of 96 possible points.

Procedures

Participants met individually with the researcher in her office. The 1.5 hours data collection session started with participants reading the consent form and filling out a language background questionnaire. They then completed the production task using Audacity and an Audio-Technica AT2020USB microphone (10 minutes), the perception tasks (AV-A-V or A-AV-V; 9 minutes per modality), lipreading tasks that are not discussed in the current study (10 minutes), the PSTM task (5 minutes), and the WM task (20 minutes). For every task involving listening, stimuli were presented via high quality Sennheiser HD380pro headphones. Breaks were provided between each task and at regular intervals within longer tasks to limit the effect of fatigue.

RESULTS

The first goal of the study was to investigate how Australian English native speakers perceive and produce the three Parisian French nasal vowels. Participants' performance on the perception task are illustrated in Figure 1. In the three modalities of presentation (AV-A-V), results showed that [ɔ̃] was the best and [ɑ̃] the least well perceived vowels. A repeated measures ANOVA with modality and vowel as within-subject factors revealed significant main effects of modality (F (2, 62) = 8.569, p = .001, η^2 = .217) and vowel (F (1.62, 50.30) = 17.333, p < .001, η^2 =.359), but no significant vowel × modality interaction (F (3.28, 101.84) = 1.552, p = .202, η^2 = .048). Bonferroni pairwise comparisons indicated no significant differences in vowel perception between the AV and A modalities (p = .404), but significant differences between the A and V modalities (p = .002) and close to significant differences in the AV and V modalities (p = .054). In terms of vowels, participants were significantly better at perceiving [ɔ̃] than [ɑ̃] (p < .001) and [ɛ̃] (p = .005), but there was no statistical difference between [ɑ̃] and [ɛ̃] (p = .247).

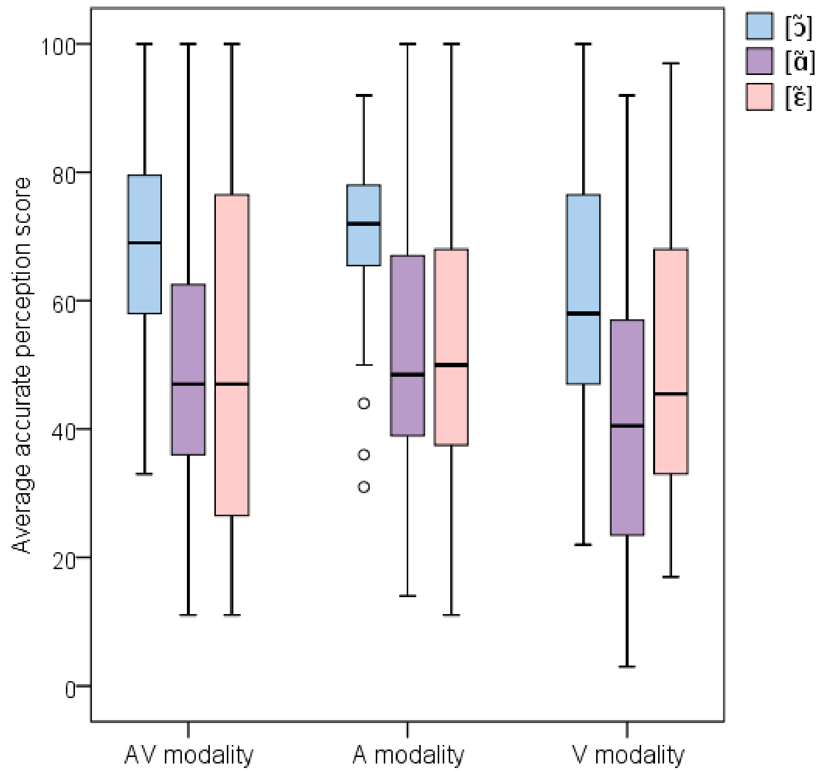


Figure 1. Mean percentage of correct perception scores for each nasal vowel in the three modalities of presentation (AV-A-V).

Results from a repeated measures ANOVA showed that there were significant differences in accurate production of the three vowels (F (1.579, 48.945) = 45.643, p < .001, η^2 =.596, *power* = 1.000) with follow-up Bonferroni pairwise comparisons indicating significant differences between [ɛ̃] and both [ɑ̃] and [ɔ̃] (p < .001).

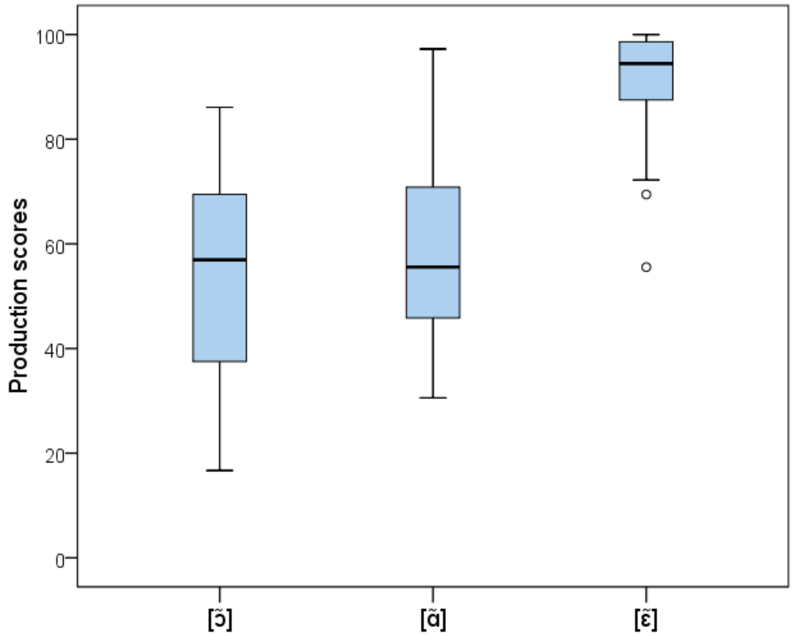


Figure 2. Mean percentage of correct production scores for each nasal vowel.

The second goal of the current study was to explore the relationship between working memory, phonological memory and speech perception. The participants' performance at the verbal WM span test are presented in Table 1, and the mean for the nonword repetition task measuring PSTM was 17.31 (out of 32) with scores ranging from 10 to 24 (standard deviation = 3.97). Results from a bivariate correlation indicated that participants' scores on these two tasks were not related (r = .227, p = .212).

Table 1

Scores for the verbal working memory span test

	Plausibility (max 24)	Grammaticality (max 24)	Recall (max 48)	TOTAL (max 96)
Mean	20.53	19.00	36.28	75.81
SD	1.56	1.83	5.87	6.91
Maximum	23	21.5	48	92
Minimum	15.5	14.5	25	60.5

In addition, participants' production scores were strongly correlated with their perception scores in the three modalities of presentation (AV: r = .66, p < .001; A: r = .65, p < .001; V: r = .68, p < .001). A set of simple linear regression models was performed to examine the extent to which WM and PSTM were predictive of L2 speech perception and production. The results showed a significant, positive relationship of moderate strength between speech perception and PSTM, indicating that learners with higher PSTM scores identified L2 vowels better. This association was significant in all modalities of presentation and indicated that 14% (AV), 21% (A), and 20% (V) of the variance in perception scores could be explained by PSTM. The analysis also confirmed a significant (p = .01) positive relationship between speech production and PSTM, whereby PSTM explained 17% of the variance. However, no relationship between WM and L2 phonology (i.e., speech perception and production) was observed.

DISCUSSION

The first goal of this study was to examine the perception and production of French nasal vowels by Australian English speakers. Results showed that [ɛ̃] was the most accurately produced vowel, with a very high score of 91%, whereas [ɑ̃] (60%) and [ɔ̃] (55%) were less accurately pronounced. These findings are very much in line with a previous study that used the same stimuli and procedures as the current study but was conducted at a US Midwestern university (accuracy scores: [ɛ̃] 78%, [ɔ̃] 61%, and [ɑ̃] 57%) (Inceoglu, 2016). However, in a study with five L1 Japanese and five L1 Spanish high intermediate learners of French, Detey and his colleagues (2010) found that [ɔ̃] was produced more accurately (67%) than [ɑ̃] (54%) and [ɛ̃] (51%). One way of accounting for these differences lies in the two methodologies used to collect learners' production data. On the one hand, the stimuli for the delayed repetition task used in the current study and Inceoglu's (2016) study consisted of 108 CVC tokens in a variety of consonantal contexts. On the other hand, Detey and colleagues (2010) used nine real words in a repetition task and a reading task, raising the issue of lexical familiarity as there is abundant evidence that lexical knowledge influences how L1 and L2 speakers perceive or recognize words (Bundgaard-Nielsen, Best, & Tyler, 2011; Flege, Takagi, & Mann, 1996). Another possible explanation that would need to be further investigated with a larger sample of participants is the L1 background of the learners. In terms of perception, the results of the current study are in line with Inceoglu (2016) showing that [ɔ̃] is significantly more accurately perceived than both [ɑ̃] and [ɛ̃] regardless of the modality of presentation.

The second and main research question explored how WM and PSTM were related to speech perception and production. First of all, the lack of correlation between the two memory tasks provided support for the assumption that PSTM and WM capacity are distinct constructs, as noted by previous studies in other areas of second language acquisition (Gathercole, 2006; Kormos & Sáfár, 2008). Importantly, the current findings revealed that achievement in the L1 PSTM task (i.e., non-word repetition task) was a good predictor for success in L2 speech perception and production. This expanded the important role of PSTM already observed in (L2) vocabulary acquisition (e.g., Speciale et al., 2004), grammar learning (e.g., Ellis & Sinclair, 1996; Williams & Lovatt, 2003) and fluency (Kormos & Sáfár, 2008). Despite differences in target languages and procedures, the current results are in line with previous studies examining speech perception (Cerviño-Povedano & Mora, 2011), pronunciation ratings (Kondo, 2011; Nagle, 2013), and vowel quality production (Zahler & Lord, 2018), and confirmed that PSTM plays a role in the acquisition of L2 speech perception and production. In terms of pedagogical implications, language learners with lower PSTM would benefit from tasks relying on repetition (i.e., activation of the phonological loop) and promoting automatization of the L2 system (Trofimovich & Gatbonton, 2006), which would allow PSTM to be redeployed for the development of long-term memory representations of L2 sounds. Finally, the lack of predictive effect of WM is similar to what Posedel et al. (2012) reported in their investigation of L2 pronunciation development, but differ from Darcy et al.'s (2005) analyses of phonological processing tasks, possibly due to differences in tasks and measures of WM.

To conclude, the current study provided interesting insights into the factors that contribute to successful L2 speech perception and production, and is one of the very few studies that aimed to explore the association between WM, PSTM and speech perception/production. Nevertheless, it is important to stress that research in this domain is still scarce and future work is needed. In particular, future studies should explore the combination effect of proficiency and should expand the investigation to other L1/L2 groups.

ABOUT THE AUTHOR

Solène Inceoglu is Lecturer (Assistant Professor) in French in the School of Literature, Languages, and Linguistics at the Australian National University. She received her Ph.D. in Second Language Studies from Michigan State University. Her research focuses on second language acquisition, second language speech perception/production, pronunciation instruction, and psycholinguistics. Contact information: solene.inceoglu@anu.edu.au

REFERENCES

Aliaga-García, C., Mora, J. C., & Cerviño-Povedano, E. (2011). Phonological short-term memory and L2 speech learning in adulthood. *Poznań Studies in Contemporary Linguistics, 47*, 1–14.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*, 417–423.

Baddeley, A., & Hitch, G. (1974). Working memory. In G. A. Bower (Ed.), *The psychology of learning and motivation* (pp. 47–89). New York, NY: Academic Press.

Bundgaard-Nielsen, R. L., Best, C. T., & Tyler, M. D. (2011). Vocabulary size matters: The assimilation of second-language Australian English vowels to first-language Japanese vowel categories. *Applied Psycholinguistics*, *32*, 51–67.

Cedrus. (2015). SuperLab 5.0 Cedrus Corporation.

Cerviño-Povedano, E., & Mora, J. C. (2010). Investigating Catalan-Spanish bilingual EFL learners' over-reliance on duration: Vowel cue weighting and phonological short-term memory. In K. Dziubalska-Kołaczyk, M. Wrembel, & M. Kul (Eds.), *Achievements and perspectives in the acquisition of second language speech: New sounds 2010* (pp. 83–88). Frankfurt am Main: Peter Lang.

Coughlin, C. E., & Tremblay, A. (2013). Proficiency and working memory based explanations for nonnative speakers' sensitivity to agreement in sentence processing. *Applied Psycholinguistics*, *34*, 615–646.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory during reading. *Journal Of Verbal Learning And Verbal Behavior*, *19*, 450–466.

Darcy, I., & Mora, J. C. (2015). Tongue movement in a second language" The case of Spanish /ei/-/e/ for English learners of Spanish. In *International Congress of Phonetic Sciences*.

Detey, S., Racine, I., Kawaguchi, Y., Zay, F., Buehler, N., & Schwab, S. (2010). Evaluation des voyelles nasales en français L2 en production : De la nécessité d'un corpus multitâches. In F. Neveu, V. Muni-Toké, J. Durand, T. Klingler, L. Mondada, & P. S (Eds.), *2ème Congrès Mondial de Linguistique Française* (pp. 1289–1301). Paris: ILF.

Ellis, N. C., & Sinclair, S. G. (1996). Working Memory in the Acquisition of Vocabulary and Syntax: Putting Language in Good Order. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *49*, 234–250.

Flege, J. E., Takagi, N., & Mann, V. (1996). Lexical familiarity and English-language experience affect Japanese adults' perception of /ɹ/ and /l/. *The Journal of the Acoustical Society of America*, *99*, 1161–1173.

French, L. M. (2006). *Phonological working memory and L2 acquisition: A developmental study of Quebec francophone children learning English*. New York, NY: Edwin Mellen Press.

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics*, *27*, 513–543.

Grey, S., Williams, J. N., & Rebuschat, P. (2015). Individual differences in incidental language learning: Phonological working memory, learning styles, and personality. *Learning and Individual Differences*, *38*, 44–53.

Harrington, M., & Sawyer, M. (1992). L2 working memory capacity and L2 reading skill. *Studies in Second Language Acquisition*, *14*, 25–38.

Inceoglu, S. (2015). Audiovisual and auditory-only perceptual training: Effects on the pronunciation of French nasal vowels. In J. M. Levis, M. Qian, & Z. Zhou (Eds.), *Proceedings of the 6th Pronunciation in Second Language Learning and Teaching Conference* (pp. 104–114). Ames, IA: Iowa State University.

Inceoglu, S. (2016). Effects of perceptual training on second language vowel perception and production. *Applied Psycholinguistics*, *37*, 1175–1199.

Kissling, E. M. (2014). What predicts the effectiveness of foreign-language pronunciation instruction? Investigating the role of perception and other individual differences. *Canadian Modern Language Review*, *70*, 532–558.

Kondo, A. (2012). Phonological memory and L2 pronunciation skills. In A. Stewart & N. Sonda (Eds.), *JALT 2011 Conference Proceedings* (pp. 535–541).

Kondo, A. (2018). Effects of Japanese EFL learners' acoustic short-term memory on English word reproduction skills. Paper presented at the 10th Pronunciation in Second Language

Learning and Teaching Conference. Iowa State University: Ames, Iowa.

Kormos, J., & Sáfár, A. (2008). Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and Cognition, 11*, 261–271.

Kroll, J. F., Michael, E., Tokowicz, N., & Dufour, R. (2002). The development of lexical fluency in a second language. *Second Language Research*, *18*, 137–171.

Lado, B. (2008). *The role of bilingualism, type of feedback, and cognitive capacity in the acquisition of non-primary languages: a computer-based study*. Washington D.C: Unpublished doctoral dissertation, Georgetown University.

Lengeris, A., & Nicolaidis, K. (2014). English consonant confusions by Greek listeners in quiet and noise and the role of phonological short-term memory. In *Proceedings of the 15th Annual Conference of the International Speech Communication Association* (pp. 534–538).

MacKay, I. R. A., Meador, D., & Flege, J. E. (2001). The identification of English consonants by native speakers of Italian. *Phonetica, 58*, 103–125.

Mota, M. (2003). Working memory capacity and fluency, accuracy, complexity, and lexical density in L2 speech production. *Fragmentos, Revista de Língua e Literatura Estrangeiras*, *24*, 69–104.

Nagle, C. (2013). A reexamination of ultimate attainment in L2 phonology: Length of immersion, motivation, and phonological short-term memory. In E. Voss, S.-J. D. Tai, & Z. Li (Eds.), *Selected Proceedings of the 2011 Second Language Research Forum* (pp. 148–161). Sommerville, MA: Cascadilla Proceedings Project.

O'Brien, I., Segalowitz, N., Collentine, J. O. E., & Freed, B. (2006). Phonological memory and lexical, narrative, and grammatical skills in second language oral production by adult learners. *Applied Psycholinguistics*, *27*, 377–402.

O'Brien, I., Segalowitz, N., Freed, B., & Collentine, J. (2007). Phonological memory predicts second language oral fluency gains in adults. *Studies in Second Language Acquisition*, *29*, 557–581.

Posedel, J., Emery, L., Souza, B., & Fountain, C. (2012). Pitch perception, working memory, and second-language phonological production. *Psychology of Music*, *40*, 508–517.

Safronova, E., & Mora, J. C. (2012). Acoustic and phonological memory in L2 vowel perception. In *Proceedings of 22nd EUROSLA* (pp. 384–390).

Serafini, E. J., & Sanz, C. (2016). Evidence for the decreasing impact of cognitive ability on second language development as proficiency increases. *Studies in Second Language Acquisition*, *38*, 607–646.

Speciale, G., Ellis, N. C., & Bywater, T. (2004). Phonological sequence learning and short-term store capacity determine second language vocabulary acquisition. *Applied Psycholinguistics*, *25*, 293–321.

Trofimovich, P., & Gatbonton, E. (2006). Repetition and focus on form in learning L2 pronunciation: Using insights from speech processing to inform L2 pronunciation teaching. *Modern Language Journal*, *90*, 519–535.

Williams, J. N., & Lovatt, P. (2003). Phonologiocal memory and rule learning. *Language Learning*, *53*, 67–121.

Winke, P. (2013). An investigation into second language aptitude for advanced Chinese language learning. *The Modern Language Journal*, *97*, 109–130.

Zahler, S., & Lord, G. (2018). The role of phonological short-term memory in Spanish second language phonology: Exploring vowel quality and duration among English-speaking learners. Paper presented at the 10th Current Approaches to Spanish and Portuguese Second Language Phonology, Indiana University: Bloomington, IN.

Koffi, E. (2019). A template model account of lexical stress in Arabic-accented English. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Ames, IA, September 2018 (pp. 158-167). Ames, IA: Iowa State University.

PRESENTATION/POSTER

A TEMPLATE MODEL ACCOUNT OF LEXICAL STRESS IN ARABIC-ACCENTED ENGLISH

Ettien Koffi, St. Cloud State University

This study applies Fry's (1958) seminal methodology to investigate how Arabic speakers of English encode and rank F0, intensity, and duration in their pronunciation of disyllabic words. At issue is whether or not Arabic speakers transfer the acoustic correlates of word stress from their L1 into their L2. Al-Ani (1992) found that Arabic speakers encode lexical stress in their L1 by relying overwhelmingly on intensity. Bouchhioua (2008) noted that Tunisian Arabic speakers relied on duration to encode word stress in their L2 English. We explore the issue further by analyzing the pronunciation of 10 Arabic speakers of English who read the Speech Accent Archive text containing seven disyllabic words.

INTRODUCTION

The importance of lexical stress in the intelligibility of L2-accented English is overstated. Confusion persists because the assessment of the role that suprasegmentals play in intelligibility has been largely impressionistic. In this paper, we propose an acoustic phonetic methodology that helps to gauge the role of suprasegmentals accurately. The paper is divided into five main sections. The first provides a quick overview of suprasegmentals in both English and Arabic. The second highlights the main findings on the acoustic correlates of lexical stress. The third gives an overview of the Template Model (TM). The fourth proceeds with the acoustic phonetic measurements of the data. The fifth examines pedagogical implications and applications.

Brief overview of suprasegmentals in English and Arabic

English is classified as a stress-timed language, that is, "some syllables will be longer and some shorter and the intervals between stressed syllables are roughly of equal length" (Fromkin et al. 2017, p. 205). Astruc (2013) and Dehmam and Lobeck (2013) classify Arabic as a stress-timed language. In Arabic and English, syllable-weight determines the placement of lexical stress. Goldsmith (1990) indicates that in Arabic, primary stress falls mostly on super heavy codas, otherwise on penultimate syllables. The same quantity-sensitive stress rule applies to disyllabic words in English. We limit our inquiry to disyllabic words because, according to Lehiste (1970), they are the minimal units where differential stress patterns can be optimally observed. A study of 190,000 English words by Cutler and Carter (1987) reveals that 39% of them are disyllabic. Of these, 90% have primary stress on the penultimate syllable. In other words, disyllabic words overwhelmingly have trochaic feet. Only 10% of English disyllabic words have an iambic foot. For this reason, Chrabaszcz et al. (2014) contend that English speakers have a trochaic bias. Words with trochaic patterns are also commonly found in Arabic (Kenstowicz, 1994). Yet, it remains to be seen whether or not the two languages rely on the same correlate ranking strategy to encode lexical stress.

Ranking of the acoustic correlates of lexical stress in English

The research on the acoustic correlates of lexical stress and their ranking was pioneered by Fry. In three seminal papers, 1955, 1958, and 1965, he found that native speakers of English rely on the following rankings: **Duration > Intensity** (1955), **F0 > Duration > Intensity** (1958), **F0 > Duration > Intensity > F1** (1965). The 1958 ranking is the most widely known and cited. A correlate ranking war of some sort has since ensued. Replication studies have come up with different rankings, including: **Duration > F0 > Intensity**; **Intensity > Duration > F0**; or **F0 > intensity > Duration**, etc. (see Koffi 2018b, pp. 15-16 for an extensive review of ranking and counter-ranking proposals). The ranking of correlates is important because it helps to determine the prosodic strategy that L2 speakers of English of the same L1 background are most likely to rely on in encoding lexical stress. Furthermore, it helps answer the question of whether or not different ranking strategies hinder or facilitate suprasegmental intelligibility.

Correlate ranking in Arabic

Al-Ani (1992), de Jong and Zawaydeh (1999), and Bouchhioua (2008) contain information about the acoustics of lexical stress in Arabic and some ranking of correlates. Al-Ani (1992) found that when speaking in Arabic, Saudi speakers encode lexical stress and rank their correlates as follows: **Intensity > Duration > F0**. de Jong and Zawaydeh (1999) did not rank F0, intensity, and duration but noted quite clearly that duration played an extremely important role in lexical stress in Jordanian Arabic. Bouchhioua (2008) studied Arabic-accented English words produced by Tunisia speakers and found that the participants encoded lexical stress and ranked correlates as **Duration > Intensity > F0**. It is noteworthy that in all these studies, F0 ranks lower than intensity and duration.

Overview of the Template Model

Various models of auditory perception of speech exist. Some are discussed in Massaro and Jesse (2005). The model used in this paper is based on Rabiner's (1999) Template Model (TM) because it makes it possible to assess intelligibility instrumentally instead of doing so impressionistically. An impressionistic assessment of relies on human judges' opinions to determine whether or not a segment or a suprasegment is intelligible. It is by far the most commonly used methodology in pronunciation research, but it is not necessarily the most accurate. The instrumental methodology, on the other hand, gauges intelligibility by measuring the frequency, intensity, and duration imbedded in the speech signals emitted by the talker. Physicist Harvey Fletcher, the inventor of the modern audiogram machine, pioneered this approach (Yost, 2015). In a seminal paper in 1940, he calculated mathematically the frequency responses of speech signals in the basilar membrane. This is now known in acoustic phonetics circle as the Critical Band Theory (CBT). Another physicist, von Bekesy, spent 20 years verifying and confirming clinically that Fletcher's calculations were grounded in physiological reality. For this, Bekesy was awarded the Nobel Prize in Physiology/Medicine in 1961. CBT thresholds have undergone some refinements since then and are now endorsed by the American National Standards Institute (ANSI), the International Electrotechnical Commission (IEC), and other reputable regulatory bodies for the manufacturing of audio products and sound level meters (Pope, 1998). Researchers in a wide variety of fields, audiology, acoustics, automatic speech recognition, speech digitalization, speech synthesis, etc.

rely on CBT thresholds to assess or simulate speech intelligibility. This paper applies relevant CBT-derived thresholds to account for the intelligibility of lexical stress in Arabic-accented English. Proponents of CBT hold that the speech signals emitted by the talkers retain their essential acoustic phonetic properties in the basilar membrane and well into the Central Auditory Nervous System (Yost, 2007, pp. 223-248).

According to TM, for automatic speech recognition by humans or machines, all one needs is a simple algorithm that calculates arithmetic means and standard deviations. A word is automatically recognized if it deviates from the exemplar within acceptable limits of the standard deviation. When TM is applied to the intelligibility of suprasegmentals, we deduce that the closer the acoustic correlates produced by the talker match the exemplar in the mind of the hearer, the more felicitous the recognition. The application of TM calls for knowing the suprasegmental characteristics of lexical exemplars in the hearer's phonological memory. Many linguists, including Fromkin et al. (2017), have given us some clues concerning these suprasegmental characteristics, saying, "In many languages, certain syllables in a word are louder, slightly higher in pitch, and somewhat longer in duration than other syllables in the word. They are stressed syllables" (p. 205).

In order to demonstrate this mathematically, we must first convert impressionistic terms such as "louder," "higher in pitch," and "longer in duration" into measurable and quantifiable correlates. Fortunately, nearly 100 years of psychoacoustic research have made the conversion possible. Important Just Noticeable Thresholds (JNDs) of pitch, duration, and intensity have been discovered which allow us to translate the above-mentioned impressionistic terms into measurable entities. The suprasegmental characteristics of lexical stress can now be restated mathematically as follows:

> A strong/stressed syllable is one whose F0 is ≥1 Hz higher, whose intensity is ≥3 dB louder, or whose duration is ≥10 ms longer than any other syllable(s) within the same word.

The list of authorities who have discussed these JNDs is long and impressive. Suffice it to mention only Lehiste (1970) for the JND in pitch, Moore (2007) for the JND in intensity, and Hirsh (1959) for the JND in duration. These authoritative JNDs are also summarized and discussed in Stevens (2000) and Yost (2007). In the remainder of the paper, we use these JNDs in tandem with TM to account for the intelligibility of lexical stress in Arabic-accented English.

DATA ANALYSIS, PARTICIPANTS, AND ANNOTATION PROCEDURES

The preceding sections have provided the background. Let's now apply these insights to examine how the 10 participants in our study encode lexical stress on disyllabic English words. The corpus comes from George Mason University's Speech Accent Archive (http://accent.gmu.edu/howto.php) and contains relevant sociometric information about the participants. The text contains 69 words, seven of which are disyllabic (10.14%). The disyllabic words and their corresponding IPA transcription are in Table 1. Stressed syllables are highlighted in bold both in spelling and in IPA transcription.

Table 1

IPA transcription of disyllabic words

N0	Word	IPA
1.	St**é**lla	[ˈst**ɛ**lə]
2.	m**á**ybe	[ˈm**e**bi]
3.	br**ó**ther	[ˈbr**ʌ**ðər]
4.	**á**lso	[ˈ**ɔ**:lso]
5.	pl**á**stic	[ˈpl**æ**stɪk]
6.	W**é**dnesday	[ˈw**ɛ̃**nzde]
7.	st**á**tion	[ˈst**e**ʃn̩]

The transcription of lexical stress is based on *Oxford Advanced Learner's Dictionary* (OALD 2000). The nucleus of each syllable is measured according to F0, intensity, and duration, as shown in the annotation in Figure 1 below. The total number of tokens examined in this study is 420 (7 words x 2 syllables x 10 participants).

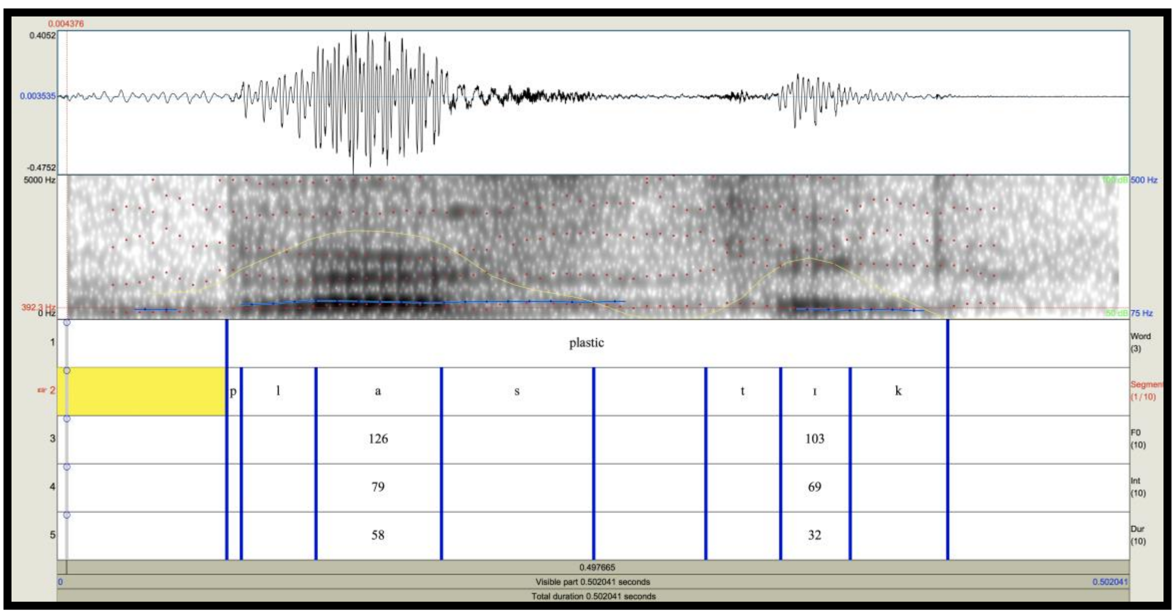


Figure 1. Annotation of "Plastic" by Arabic 36M.

This annotation shows that only the nuclei <a> in <plas> and <i> in <tic> are measured. The same procedure is repeated for all the seven disyllabic words in the data.

Acoustic Measurements of Suprasegments in Arabic-Accented English

The JND thresholds mentioned previously are now applied to account for how the 10 participants encode and rank the acoustic correlates of lexical stress in Arabic-accented English. Tables 2, 3, and 4 display measurements for F0, intensity, and duration. Unless otherwise stated, the analyses focus on the arithmetic means, not on idiosyncratic pronunciations of individual words or participants.

Table 2

F0 measurements

Words	St**é**lla		m**áy**be		br**ó**ther		**á**lso		pl**á**stic		W**é**dnesday			st**á**tion	
F0	ste	la	may	be	bro	ther	al	so	plas	tic	we	nis	day	sta	tion
Arabic 1F	247	213	254	268	224	256	241	252	229	281	247	257	276	226	208
Arabic 30F	225	204	218	223	218	216	236	286	229	273	220		207	74	275
Arabic 35M	128	145	125	124	114	111	123	114	116	74	108	117	119	84	74
Arabic 36M	128	98	153	98	120	118	142	130	126	103	123		120	105	74
Arabic 40M	116	115	136	120	124	127	154	154	128	112	124		119	101	95
Arabic 44F	187	188	224	248	212	223	222	271	201	156	224		192	149	74
Arabic 46M	103	110	126	136	125	137	128	131	124	141	127	126	127	74	145
Arabic 47M	113	147	117	113	113	94	171	116	113	115	121	128	125	113	92
Arabic 50M	101	108	98	94	84	96	112	111	95	125	102	99	96	88	74
Arabic 51M	119	111	121	122	107	108	129	136	121	116	109		108	75	117
Arabic Mean	**146**	**143**	**157**	**154**	**144**	**148**	**165**	**170**	**148**	**149**	**150**	**145**	**148**	**108**	**122**
St. Dev.	53	43	54	65	52	59	49	70	50	70	56	8	57	47	68

Table 3

Intensity measurements

Words	St**é**lla		m**áy**be		br**ó**ther		**á**lso		pl**á**stic		W**é**dnesday			st**á**tion	
Intensity	ste	la	may	be	bro	ther	al	so	plas	tic	we	nis	day	sta	tion
Arabic 1F	80	72	81	81	78	77	74	78	79	82	82	80	77	77	72
Arabic 30F	77	74	80	75	82	82	73	78	75	73	78		75	68	64
Arabic 35M	72	76	73	69	76	76	76	72	76	63	74	73	75	66	56
Arabic 36M	76	69	79	69	78	79	83	75	79	69	75		77	71	59
Arabic 40M	75	71	75	76	75	71	80	76	77	72	74		76	69	63
Arabic 44F	75	74	82	72	78	85	79	75	76	73	82		78	76	68
Arabic 46M	73	71	77	75	81	81	83	75	73	73	77	76	76	69	63
Arabic 47M	81	74	81	75	83	77	81	79	80	79	81	81	80	84	70
Arabic 50M	78	80	81	80	78	76	80	75	82	74	81	80	81	71	60
Arabic 51M	74	67	77	76	74	73	79	77	77	71	72		75	66	57
Arabic Mean	**76**	**72**	**78**	**74**	**78**	**77**	**78**	**76**	**77**	**72**	**77**	**78**	**77**	**71**	**63**
St. Dev.	2	3	2	3	2	4	3	2	2	5	3	1	2	5	5

Table 4

Duration measurements

Words	St**é**lla		m**á**ybe		br**ó**ther		**á**lso		pl**á**stic		W**é**dnesday			st**á**tion	
Duration	ste	la	may	be	bro	ther	al	so	plas	tic	we	nis	day	sta	tion
Arabic 1F	85	97	225	99	115	155	234	293	149	127	99	104	241	159	89
Arabic 30F	80	77	97	96	67	33	115	59	81	50	65		74	63	59
Arabic 35M	97	103	146	160	76	60	140	62	91	30	48	60	177	94	39
Arabic 36M	73	40	128	88	110	74	112	55	58	32	88		190	64	70
Arabic 40M	76	89	164	61	72	56	170	84	93	53	59		240	100	41
Arabic 44F	67	72	104	109	55	55	140	39	64	47	112		200	65	52
Arabic 46M	91	78	153	76	50	48	38	79	75	80	77	72	219	102	71
Arabic 47M	159	40	153	125	76	80	140	153	75	157	86	64	278	84	98
Arabic 50M	73	91	84	156	74	61	123	55	51	44	89	93	175	83	57
Arabic 51M	60	95	69	82	55	50	96	75	63	41	75		133	70	88
Arabic Mean	**86**	**78**	**132**	**105**	**75**	**67**	**130**	**95**	**80**	**66**	**79**	**78**	**192**	**88**	**66**
St. Dev.	27	22	46	32	21	33	50	76	27	42	19	16	58	28	20

The JND in F0 shows that in 3 of 7 words (42.85%), the nuclei of the penultimate syllables in <Stella>, <maybe>, and <Wednesday> are at least 1 Hz higher than those of the unstressed syllables. In these cases, the stress pattern conforms to the phonological exemplar in the mental lexicon of native speakers as putatively represented by (OALD, 2000). But such is not the case for <brother>, <also>, <plastic>, and <station>. The JND in intensity indicates that in 4 out of 7 words (57.14%), i.e., <Stella>, <maybe>, <plastic>, and <station>, the nuclei of the penultimate syllables are at least 3 dB louder than the nuclei in the ultima. However, this is not so for <also>, <brother>, and <Wednesday>. The latter is a special case because half of the participants resyllabified it into three syllables. Instead of [ˈwɛ̃nz.de], it turned into [ˈwɛ̃.niz.de]. We see that the nuclei of the penultimate syllables of <maybe>, <also>, <plastic>, and <station> are at least 10ms longer than the unstressed nuclei in <Stella> and <brother>. So, 4 out 7 words (57.14%) conform to the threshold for the JND in duration.

The data in the three tables indicate that lexical stress is produced intelligibly in six words. The only exception has to do with <brother>. According to the exemplar, the F0, intensity, and duration of the vowel [ɑ] in the penultimate syllable should be at least 1 Hz higher, 3 dB louder, and 10 ms longer than the vowel [ə] in the ultima. However, only 4 out 10 participants pronounced <brother> as expected in regard to F0 and intensity. Furthermore, only 2 out 10 produced the duration as expected. Why? It all has to do with the fact that six participants trilled the post-vocalic [r] in the coda of <ther>. Trilling this [r] affects the acoustic correlates of the preceding vowel. Stevens (2000) notes that raising the blade of the tongue to trill [r] increases the F0 of the preceding vowel. In other words, the F0 of [ə] becomes higher than that of [ɑ]. Lehiste (1970) provides an articulatory rationale for why trilling [r] increases the F0 of the preceding vowel, “Now, the muscles constituting the tongue are attached to the superior part of the hyoid bone, and some of the laryngeal muscles are attached to the inferior part. When the tongue is raised, the larynx tends to be pulled upwards and the laryngeal muscles are stretched. This increases the tension of the vocal folds and causes the increase in the vibration rate” (p. 71). Since speakers of many dialects of Arabic trill their [r]s, F0 measurements of the preceding or following vowels may not always conform to the example, as is the case of <brother>.

Correlate ranking

It stems from the measurements and analyses above that the Arabic talkers in this study rely **equally** on intensity and duration to encode lexical stress in English. The ranking of their correlates is as follows: **Intensity (57.14%) = Duration (57.14%) > F0 (42.85%)**. This ranking is consistent with the results in Al-Ani (1992) who found that intensity is the main acoustic correlate of stress in Saudi Arabic and Jong and Zawaydeh (1999) who found that the Jordanian participants in their study relied primarily on duration to encode lexical stress. Our measurements and ranking are in line with both findings in that they show the participants in our study transfer intensity or duration from their L1 to encode lexical stress in their L2 English. There is nothing unusual about this finding because we know from Yost (2007) that "different auditory neurons perceive the different physical components of sounds. Some perceive frequencies, other perceive intensity, while other perceive duration" (p. 223). Since the three acoustic correlates of stress are independent of each other, speakers can use either of them to encode lexical stress. Regardless of the correlates considered, the suprasegmentals in <Stella, maybe, also, plastic, Wednesday, station> are produced and perceived intelligibly. Only the stress pattern of <brother> deviates from the expected trochaic pattern for reasons given in the preceding paragraph.

The intelligibility of suprasegmentals in Arabic-accented English

Yavaş (2011) extrapolates on the basis of syllable weight alone that Arabic speakers of English would misplace the lexical stress on <d**í**fficult>, <**é**xpert>, <n**á**rrowest>, and <**í**nstitute> and mis-stress them as <diffic**ú**lt>, <exp**é**rt>, <narrow**é**st> and <instit**ú**te> because the ultima in all these words are heavy. However, syllable weight is not the only factor nor is it even the determinative factor in assigning lexical stress. The foot structure of the word also plays an important role. De Jong and Zawaydeh (1999) state repeatedly that Arabic and English speakers have similar stress patterns. Consequently, Arabic speakers can bypass syllable weight altogether and adopt a trochaic pronunciation regardless of syllable weight. I have personally interacted with numerous Arabic students, friends, and colleagues, but I have not heard any of them misplace the stress on these words.

Misplaced lexical stress and intelligibility

Yavaş' extrapolation is reminiscent of the type of hyperbolic statements that one encounters frequently in the English L2 pronunciation literature concerning lexical stress. Field (2005) states that "research evidence suggests that suprasegmentals play a more prominent role than segmentals" (p. 402) even though his own paper shows that "incorrect misplacement of lexical stress is, relatively speaking, quite small: affecting only around 8% of content words if every word were misstressed" (p. 417). Claiming that misplacing lexical stress undermines intelligibility betrays a proper of understanding of the auditory and neural processes involved in the perception of suprasegmentals in accent and tone languages (Koffi, 2018a). English is an accent language, not a tone language. Consequently, misplacing lexical stress alone does not and cannot impede intelligibility. Intelligibility is jeopardized only if the words whose stress is mispronounced have lexical competitors, i.e., <p**é**rmit> vs. <perm**í**t>, <**ó**bject> vs. <obj**é**ct>, <c**ó**ntract> vs. <contr**á**ct>, <s**ú**bject> vs. <subj**é**ct>, <r**é**cord> vs. <rec**ó**rd>. However, the total number of such pairs is infinitesimal compared with the tens of thousands of words in English. Even so, misplacing lexical stress on such pairs does not automatically hinder intelligibility because of contextual and syntactic redundancies in spoken utterances. These redundancies allow the hearer to recover the part of speech of words and recognize them accurately. Since English is an accent language, not a tone language, misplacing lexical stress alone cannot and does not interfere with intelligibility. If English were a tonal language, intelligibility would be a different story altogether. In such languages, lexical competitors abound. There may be two, three, four, or even five lexical minimal pairs that are distinguishable from each other by pitch alone. Such is the case of <na> in Thai which has five different meaning depending on pitch fluctuations (Fromkin et al., 2017). The only way that suprasegmentals can play havoc on intelligibility in English is if one or several segments within the same words are **also** severely mispronounced beyond recognition.

PEDAGOGICAL IMPLICATIONS AND APPLICATIONS

Monosyllabic content words that make up 45% of the 190,000 items in Cutler and Carter's (1987) corpus have a predictable stress pattern because they have only strong syllables. The stress pattern of 90% of disyllabic words in their corpus is predictably trochaic because they are of Germanic origin. Arabic speakers would have no problem producing such stress patterns intelligibly. The remaining 10% have an iambic stress pattern. Halle and Chomsky (1999) posit that such words contain the etymological feature [+foreign] in their underlying representation. According to Field (2005), the best way to teach words with non-predictable stress is to draw students' attention to them during vocabulary instruction. Arabic speakers would have no problem producing such words intelligibly if their attention is drawn to their unusual lexical stress pattern.

SUMMARY

The Template Model and the JND thresholds used in this paper allow researchers to gauge the intelligibility of suprasegmentals in L2-English accurately. Since F0, intensity, and duration are three independent and interdependent acoustic correlates, each one by itself or in tandem with others can be used to encode lexical stress. Given that these three acoustic cues are universal and can be combined in various ways to encode lexical stress in English, it is very unlikely that misplacing any one of them can impair intelligibility. This, however, does not mean that lexical stress is not important. It is, but its weight on intelligibility has been exaggerated in complete disregard of the fact that English is an accent language. If it were a tone language like many African languages, Mandarin, or Thai, it would not be an exaggeration to say that misplacing the acoustic correlates of suprasegmentals would lead to frequent communication breakdowns, and/or to occasional embarrassments.

ABOUT THE AUTHOR

Ettien Koffi, Ph.D., is a professor of Linguistics at St. Cloud State University, Minnesota, USA, specializing in acoustic phonetics (Speech Intelligibility). His research interests center around sociophonetic variations in Central Minnesota English, acoustic phonetic accounts of intelligibility in L2 English, and acoustic phonetic and general description of Anyi, a West African Language

spoken in Cote d'Ivoire. He is the author of four books and numerous papers covering topics as varied as syntax, translation, language planning and policy, orthography, and indigenous literacy training manuals. He can be reached at enkoffi@stcloudstate.edu.

REFERENCES

Astruct, L. (2013). Prosody. M. J. Jones & R.A. Knight (Eds.), *The Bloomsbury companion to phonetics* (pp. 126-139). New York: Bloomsbury.

Al-Ani, S. H. (1992). Stress variation of the construct phrase in Arabic: Spectrographic analysis. *Anthropological Linguistics 34*(1), 256-276.

Bouchhiou, N. (2008). Duration as a cue to stress and accent in Tunisian Arabic, Native English, and L2 English. Speech Prosody, Campinas, Brazil. *ISCA Archive*, pp. 535-538.

Chomsky, N., & Morris, H. (1990). *The sound pattern of English*. Cambridge, MA: The MIT Press.

Chrabaszcz, A., Winn, M., Lin, C. Y., & Idsardi, W. J. (2014). Acoustic cues to perception of word stress by English, Mandarin, and Russian Speakers. *Journal of Speech, Language, and Hearing Research, 57*(4), 1468-1479.

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech & Language, 2*(3-4), 133-142.

Denham, K., & Lobeck A. (2013). *Linguistics for everyone: An introduction*. Boston, MA: Wadsworth-Cengage Learning.

Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly, 39*(3), 399-423.

Fletcher, H. (1940). Auditory Patterns. *Reviews of Modern Physics, 12*, 47-65.

Fromkin, V., Rodman, R., & Hyams, N. (2017). *An introduction to language* (11th ed.). Boston, MA: Cengage Learning.

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America, 27*(4), 765-768.

Fry, D. B. (1976). Introduction. In D. B. Fry (Ed.), *Acoustic phonetics: A course of basic readings* (pp. 11-17). New York: Cambridge University Press.

Goldsmith, J. (1990). *Autosegmental and metrical phonology*. Cambridge, MA: Basil Blackwell.

Kenstowicz, M. (1994). *Phonology in generative grammar*. Cambridge, MA: Blackwell Publishers.

Koffi, E. (2018a). A just noticeable difference (JND) reanalysis of Fry's original acoustic correlates of stress in American English. *Linguistic Portfolios*, *7*, 2-25.

Koffi, E. (2018b). Differential analysis of lexical pitch in accent and tone languages. *Linguistic Portfolios*, *7*, 110-131.

Ladefoged, P. (2001). *A course in phonetics* (4th ed.). New York: Harcourt College Publishing.

Lehiste, I. (1976). Suprasegmental geatures in speech. In N. Lass (Ed.), *Contemporary Issues in Experimental Phonetics* (pp. 225-239). New York: Academic Press.

Massaro, D. (2005). The magic of reading: Too many influences for quick and easy explanation. In T. Trabasso, J. Sabatini, D. C. Massro, & R. C. Calfee (Eds.), *From orthography to pedagogy: Essays in honor of Richard L. Venezky* (pp. 37-61). Mahwah, NJ: Lawrence Erlbaum Associate Publishers.

Pope, J. (1998). Analyzers. In M. J. Crocker (Ed.), *Handbook of acoustics* (pp. 1341-1353). New York: John Wiley and Sons Inc.

Rabiner, L. (1998). Machine recognition of speech. In M. J. Crocker (Ed.), *Handbook of acoustics* (pp. 1263-1270). New York: John Wiley and Sons Inc.

Stevens, K. (2000). *Acoustic phonetics*. Cambridge, MA: The MIT Press.

Yost, W. A. (2015). Psychoacoustics: A brief historical overview. *Acoustics Today, 11*(3), 46-53.

Yost, W. A. (2007). *Fundamentals of hearing: An introduction*. (5th ed.). New York: Elsevier.

Lee, J., Kim, D. J., & Park, H. (2019). Native listener's evaluations of pleasantness, foreign accent, comprehensibility, and fluency in the speech of accented talkers. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Ames, IA, September 2018 (pp. 168-178). Ames, IA: Iowa State University.

PRESENTATION/POSTER

NATIVE LISTENERS' EVALUATIONS OF PLEASANTNESS, FOREIGN ACCENT, COMPREHENSIBILITY, AND FLUENCY IN THE SPEECH OF ACCENTED TALKERS

Jieun Lee, University of Wisconsin-Milwaukee
Dong Jin Kim, University of Wisconsin-Milwaukee
Hanyong Park, University of Wisconsin-Milwaukee

Compared to studies on accentedness, fluency, and comprehensibility, there are few studies on pleasantness in second language (L2) research. To address this gap, we investigated native English speakers' subjective evaluations of pleasantness, accentedness, comprehensibility, and fluency in the speech of Korean learners of English. Twenty-six raters made judgements on a 9-point Likert scale after listening to spontaneous speech samples from Korean learners of English. Results indicated that pleasantness ratings are predicted by all dimensions. In particular, we observed fluency as the best predictor of native listeners' perceived pleasantness, followed by comprehensibility and accentedness. Our findings suggest native speakers' (NSs) appraisals of L2 speech pleasantness is influenced by how fluent and comprehensible L2 speakers are.

INTRODUCTION

Listeners automatically evaluate many aspects of the speech of their interlocutors. The degree of foreign accent, comprehensibility, fluency, and intelligibility are some aspects that listeners continuously assess while conversing with non-native speakers (NNSs). Previous research has examined the relationships between these dimensions in NSs' perception of utterances by L2 learners (e.g., Derwing, Rossiter, Munro, & Thomson, 2004; Munro & Derwing, 1995a, 1999; Trofimovich & Isaacs, 2012). Pleasantness, listeners' subjective evaluation of their holistic conversational experience, may be another important dimension in L2 learning in that pleasant speech may result in successful encounters with NSs and this may increase learners' motivation to seek out more interactions. In spite of the potentially important role of pleasantness in L2 learning, there are few studies on pleasantness. In particular, questions regarding native listeners' reactions to L2 speakers' pleasantness and to what extent this dimension is related to other speech dimensions are mostly unanswered. Thus, we investigate these questions in the current study.

Research on accentedness, comprehensibility, and fluency

The overall conclusion about the relationships between accentedness, comprehensibility, and fluency is that these dimensions are associated with each other with varying degrees of strength. For example, Derwing, Rossiter, Munro, and Thomson (2004) found a stronger relationship between fluency and comprehensibility than between fluency and accentedness in NSs' ratings of Mandarin learners of English. Their results suggest that a strong foreign accent does not necessarily result in reduced fluency or comprehensibility. However, to the best of our knowledge, there is no research that attempts to examine how pleasantness constructs relationships with other

L2 speech aspects (accentedness, comprehensibility, and fluency). The major goal of the present study is to address this gap by asking native English listeners to judge the same set of stimuli on the above-mentioned four dimensions.

Research on pleasantness

Previous speech perception studies approached pleasantness as listeners' perceived (subjective) attitude toward some particular aspects of speech, such as accent, pronunciation, or voice. For instance, Giles (1970) studied listeners' perceived attitudes of the "aesthetic" content (pleasantness) of 13 different accents of native language (L1) presented both vocally and conceptually. Participants made attitude judgements on a 7-point scale (1 = extremely pleasant; 7 = extremely unpleasant). The results suggested that perceived attitudes vary for different accents, and non-linguistic factors such as age, sex, social class, and regional membership are important determinants of listeners' evaluations. Jakšič (2018) studied native Czech English as a Second Language (ESL) learners' judgements of six different varieties of English on comprehensibility, pleasantness, socioeconomic status, and model suitability. The instruction given for pleasantness ratings was "For me, the speaker's pronunciation sounds: 1 = very pleasantly [sic]; 7 = very unpleasantly [sic]" (p. 46). As in Giles (1970), listeners' judgements of pleasantness were different for several dialects.

Similar to the studies of the perceived pleasantness of speech produced by native speakers, Bouchard, Carranza, and Moffie (1977) investigated native listeners' perceived pleasantness of speech produced by L2 speakers while focusing on voices. In their study, native English listeners judged Spanish-English bilinguals' taped readings of an English passage on the likelihood of being a friend, eventual occupation, accentedness, pleasantness, and fluency using 7-point scales. Raters were told to assess each recording on the basis of voice cues alone, just as one might judge a person if the individual were talking on the telephone or speaking on the radio. The correlations among all five rating dimensions were statistically significant, and pleasantness showed the strongest positive correlation with fluency.

Along with Bouchard *et al*. (1977), which showed the close relationship between pleasantness and fluency, Derwing and Munro's (2009) study on preference suggests a possible relationship among pleasantness, comprehensibility, and fluency. They examined how L2 speech comprehensibility influences native English listeners' preference for interacting with Mandarin and Slavic learners of English. After listening to a pair of extemporaneously spoken L2 speech samples, the listeners chose the sample that they preferred. They also had an opportunity to write comments about each sample they heard. Overall, they preferred more comprehensible speech regardless of the speaker's L1. It is also worth mentioning that comprehensibility-related (e.g., easy to follow) and fluency-related (e.g., broken speech) comments were a large part of comments for the selected (preferred) voices (41%) and the non-selected (non-preferred) voices (51%). Although pleasantness in the current study and preference in Derwing and Munro's (2009) study are not identical concepts, these two have something in common; both of them would be likely to lead more interaction between L2 learners and NSs.

The current study

Unlike previous studies, in the current study on the relationships between pleasantness and accentedness, comprehensibility, and fluency, we focus on English native listeners' impressions of pleasantness in a holistic way rather than asking them to pay attention to certain features of speech. We took this approach to pleasantness because we were not confident that listeners would be able to evaluate a speech dimension and focus on one aspect of speech while ignoring others. Thus, we asked the listeners to simply rate the pleasantness of each L2 speech on a 9-point Likert scale.

This study is guided by the following research question: How is pleasantness related to accentedness, comprehensibility, and fluency? Based on previous research, we predict that pleasantness is more closely aligned with fluency and comprehensibility than accentedness.

METHODS

Talkers

Twenty-one L1-Korean speakers of L2-English (15 females, 6 males) produced the stimuli for monetary compensation. They were either university students or residents living in the Midwestern USA (mean age = 27.14, *SD* = 6.21). The average length of residence (LOR) of the L2 speakers was 2.82 years (*SD* = 1.95) and the average age of arrival (AOA) was 23.90 years (*SD* = 5.77). These production data were collected as part of a larger study conducted by Darcy, Park and Yang (2015).

Listeners

Twenty-six NSs of American English (18 females and 5 males) participated in the current study as raters for course credit. They were university students in the Midwest (mean age = 25.54 years, *SD* = 7.66).

Procedures

The L2 speakers participated in a narrative retelling task of a summary of the *North Wind and the Sun.* After reading the passage on a computer screen, they retold the story. The participants were unaware that they would be retelling the story when they read the passage. Recordings were made in a sound-attenuated room with a microphone at a sampling rate of 44.1 kHz. Excluding the first and the last sentences of the utterances of each speaker, two sentence-long tokens per speaker were selected as stimuli. These criteria for token selection were to avoid speakers' initial hesitation and disfluencies which might appear due to the sudden start of the retelling task and also prevent raters from being bored due to a lengthy experiment. Although sentence-long stimuli are not commonly used in studies on L2 fluency, we have decided to use somewhat short stimuli based on previous studies (e.g., Munro & Derwing, 2001) reporting that listeners can make reliable judgements on different aspects of L2 speech after listening to sentence-long stimuli. Another thing to consider is that our study was conducted with 42 speech samples from 21 non-native talkers and had two rating sessions (see Rating on page 171). If we followed the commonly used methods such as 30 seconds-length stimuli, the experiment would have taken almost 50 minutes, which is too long for a perceptual experiment (e.g., if we give 5 seconds to evaluate two speech dimensions for each speech sample, it takes 2,940 seconds to finish the experiment without any break: 42 tokens x (30 seconds + 5 seconds) x 2 sessions = 2,940 seconds or 49 minutes).

Filled pauses were discarded and unfilled pauses (i.e., silence) of over three seconds were modified by removing the part of pause after three seconds threshold. For example, if an unfilled pause was 3.04 seconds, only three seconds of pause remained, and the excessive 0.04 seconds were removed. Considering we used sentence-length stimuli, we were worried that too many filled pauses and extremely long unfilled pauses would becloud raters' evaluations of the L2 speech dimensions of our interest, result in a very skewed distribution in fluency ratings, and consequently make it difficult to investigate the relationships of perceptual dimensions including fluency. Although we manipulated two fluency characteristics of the samples regarding pauses, it should be noted that we considered only *uh*s and *um*s as filled pauses by following Lee (2018) and did not manipulate most characteristics of fluency, such as repetition, replacements, reformulations, hesitations, and false starts. Also, such a three seconds pause manipulation was processed only for two out of forty-two stimuli (approximately 4.76 of total stimuli). After our manipulations, we still observed a wide range of fluency ratings, suggesting that there were many fluency characteristics other than the two characteristics we manipulated. Furthermore, since our main interest is not fluency itself but the relationships among four L2 perceptual dimensions, we felt that it would be fair to examine those relationships as long as raters were given the same stimuli regardless of the manipulations of stimuli. After the editing process, the average length of selected tokens was 8.70 seconds. In total, forty-two tokens of L2-English speech from twenty-one native Korean speakers were used as stimuli in the rating session.

Rating

The experiment was conducted on a Praat platform (Boersma & Weenink, 2017) with high-quality headphones at the UWM Phonetics Lab. After completing the consent form and a language background questionnaire, the raters participated in a practice session to familiarize them with the rating task. The task consisted of two parts, with a minute of mid-session break. It took about thirty minutes to complete the rating task. In each session, the participants rated two dimensions on a 9-point Likert scale after listening to each token. For example, participants rated pleasantness and accentedness for all 42 tokens in the first session, then, after the break, they rated comprehensibility and fluency for the same set with a different randomization. All possible twenty-four combinations of rating dimensions (e.g., accentedness-comprehensibility, pleasantness-fluency) were considered and the stimuli were presented in randomized sequences. Following the definitions suggested by previous studies (e.g., Derwing & Munro, 1997; Kormos & Dénes, 2004; Munro & Derwing, 1995b; Trofimovich & Isaacs, 2012; Zetterholm & Abelin, 2017), the rating dimensions were defined as follows: Pleasantness—how pleasant or unpleasant your experience of listening to the sentence is (1 = very unpleasant, 9 = very pleasant); Accentedness—how different the speaker's accent is from standard American English (1 = very strong foreign accent, 9 = no foreign accent); Comprehensibility—how easy or difficult it is to understand the sentence (1 = impossible to understand, 9 = very easy to understand); Fluency—how fluent or disfluent the speaker is (1 = very disfluent, 9 = very fluent). These definitions were given to the raters before the practice and rating sessions on a separate piece of paper and during the sessions on a computer screen. The participants were asked to use the full-scale range when making their judgements.

RESULTS

To answer the research questions regarding the relationships among dimensions, we ran a mixed-effects model with lme4 package (Bates, Maechler, Bolker, & Walker, 2015) in R (R Core Team, 2017) after excluding outliers of the rating data based on the 3*SD* threshold (Kennedy & Trofimovich, 2008). Altogether, five instances in pleasantness, two in accentedness, one in comprehensibility, and six in fluency were excluded.

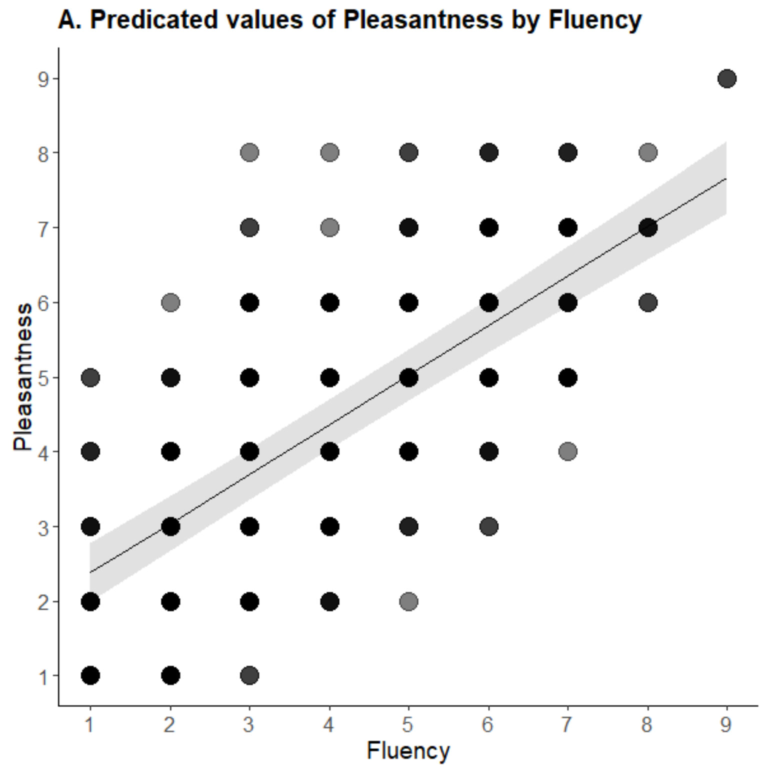
The relations among pleasantness, accentedness, comprehensibility, and fluency

Table 1 shows the linear mixed effects model of pleasantness ratings. The model contained accentedness, comprehensibility, and fluency as fixed effects. Intercepts for raters and stimuli were included as random effects. P-values and r^2 were calculated using lmerTest (Kuznetsova, Brockhoff, & Christensen, 2017) package and Nakagawa and Schielzeth approach (Nakagawa & Schielzeth, 2013) in r2glmm (Jaeger, 2017) package. The type 3 analysis of variance found that there were significant main effects of fluency (estimate = .661, *SE* = .043, t = 18.51, $F(1, 450.97)$ = 342.64, r^2 = .422, p < .001) and comprehensibility (estimate = .091, *SE* = .041, t = 2.217, $F(1, 396.06)$ = 4.92, r^2 = .015, p = .027), as well as a trend effect of accentedness (estimate = .084, *SE* = .043, t = 1.939, $F(1, 443.18)$ = 3.76, r^2 = .009, p = .053). The percentages of variance explained by the variables were 42%, 1.5%, and 0.9% respectively. Overall, more fluent, more comprehensible, and less accented L2 speech was more likely to be evaluated as being more pleasant speech. Figure 1 summarizes the results.

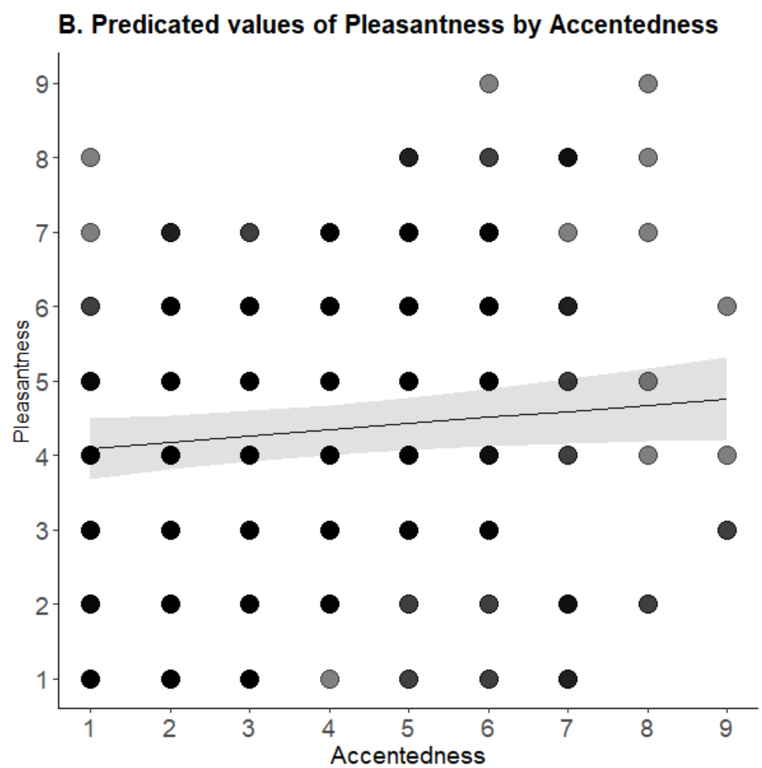
Table 1

The output of linear mixed effects regression models of pleasantness with fluency, comprehensibility and accentedness

			pleasantness		
Predictors	*Estimates*	*SE*	*t*	*Rsq*	*p*
(Intercept)	.978	.228	4.286	.547	**<.001**
Fluency	.661	.036	18.51	.422	**<.001**
Comprehensibility	.091	.041	2.217	.015	**.027**
Accentedness	.084	.043	1.939	.009	.053
Observations	455				



A. Predicated values of Pleasantness by Fluency
Pleasantness
Fluency
1
2
3
4
5
6
7
8
9



B. Predicated values of Pleasantness by Accentedness
Pleasantness
Accentedness
1
2
3
4
5
6
7
8
9

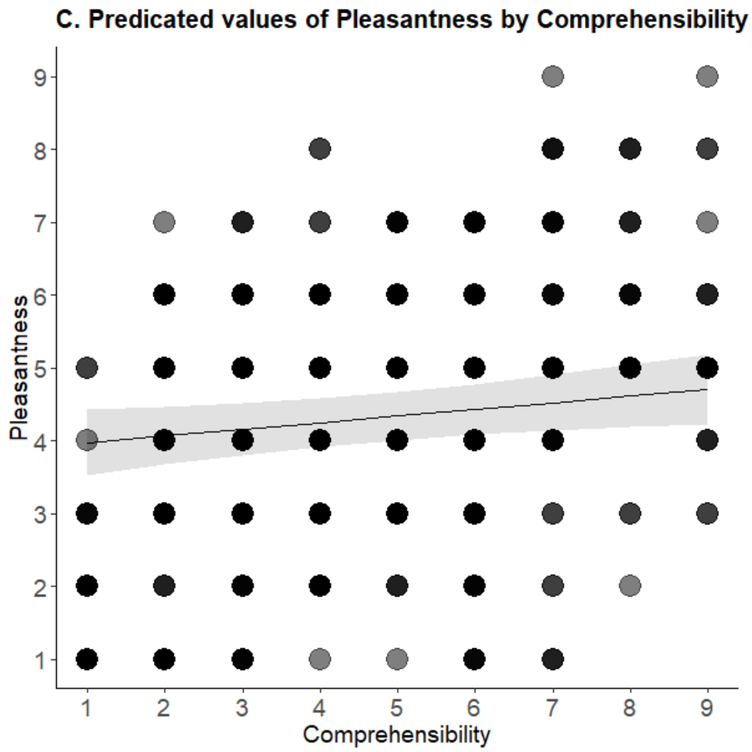


Figure 1. Prediction plots for pleasantness by fluency, accentedness, and comprehensibility. The lines show smoothed linear trends of the model-predicted effects. The shadings indicate 95% confidence-interval band. The darker dots, the more observations were made.

DISCUSSION AND CONCLUSIONS

This study investigated native listeners' evaluation of pleasantness, accentedness, comprehensibility, and fluency and their inter-relationships in L2 speech. Our study confirms previous findings that perceptual dimensions are connected to one another to varying degrees. Specifically, regarding pleasantness, our results showed that ratings of fluency seemed to be the strongest predictor of listeners' perceived pleasantness of L2 speech, followed by ratings of comprehensibility and accentedness. Our study provides support for Bouchard *et al.*'s (1977) findings. It is noteworthy that participants in their study were directed to focus on voice cues alone while rating Spanish-English bilinguals' taped readings of an English passage. It seems that in spite of listeners' special attention to bilinguals' voices, fluency still had a relatively large impact on pleasantness ratings. One possible explanation for this result is that the English passage reading stimuli in their study were quite long (forty-one words) from which listeners might use other cues, such as pronunciation errors or speech rate, along with voice cues to determine speakers' fluency levels. Adopting such long stimuli may have resulted in similar results in our study where no specific instruction was given regarding the basis of evaluation. Thus, together with Bouchard *et al.*'s (1977) findings, the results of the current study suggest that listeners' holistic impression of L2 speech pleasantness is mainly affected by how fluent L2 speakers sound.

Our predictions regarding the relationship between pleasantness, fluency, comprehensibility based on Derwing and Munro (2009) were also borne out in this study. Derwing and Munro (2009) reported that English native listeners preferred to interact with L2 speakers who had more comprehensible L2 speech. The listeners also frequently commented on comprehensibility- and fluency-related features of L2 speech when explaining their preference. Consistent with their findings, NSs' pleasantness ratings in this study were mostly predicted by fluency and comprehensibility ratings. We also noted that fluency could explain a large portion (42%) of the variance in pleasantness ratings while comprehensibility explained only 1.5% of the variance in this study. These different contributions of fluency and comprehensibility to pleasantness ratings might come from the harsher judgements on fluency than on comprehensibility by listeners in the current study (Derwing, Munro, & Thomson, 2007; O'Brien, 2014). Listeners gave the highest ratings (indicating the most positive assessment. e.g., 9 = extremely fluent) to stimuli more often for comprehensibility than for fluency or pleasantness. Thus, the variance of comprehensibility ratings was larger than the ones of fluency and pleasantness, which showed more positive skewed patterns (i.e. more frequent ratings for lower scores, which means more negative assessments). These similar patterns of pleasantness and fluency might result in the closer relationship between these two dimensions compared to comprehensibility and accentedness. In sum, fluency is proposed as a best predictor of NSs' perceived pleasantness of L2 speech in this study, although the impacts of other dimensions still exist.

A question is in order regarding the strong relationship between pleasantness and fluency: why are these dimensions closely related in evaluating L2 speech? In a follow-up study (Kim, Lee, & Park, 2018), we explored whether fluency-related linguistic properties also affected pleasantness judgements. In this vein, we measured speech rate, repair fluency, and mean length of run (Derwing et al., 2004; Kormos & Dénes, 2004; Trofimovich & Isaacs, 2012). Our results indicated that pleasantness rating scores are significantly correlated with all measures. These results suggest that the strong relationship between pleasantness and fluency ratings in the current study may come from the common linguistic properties affecting both dimensions.

Pleasantness is an important aspect of L2 speech in that more pleasant speech may increase the amount of positive verbal and non-verbal feedback from NSs, since pleasant speech may increase NSs' desire to interact with L2 learners. As a consequence, L2 learners' motivation to communicate in their target language may increase by having successful encounters with NSs. The Douglas Fir Group (2016) stated that "For L2 learners … the more they (L2 learners) experience emotionally and motivationally positive evaluations of their anticipated and real interactions, the more effort they will make to participate in them and affiliate with others" (p.28). Through this process, L2 learners are expected to build higher L2 confidence on the basis of positive L2 experiences. The importance of enhancing L2 confidence for successful L2 learning cannot be overstated (MacIntyre, 2007).

One remaining question for teachers is what might the positive pedagogical strategies for improving pleasantness be? Our study does not have a direct pedagogical component. However, based on our findings, we propose that utilizing existing pedagogies for fluency or comprehensibility improvement may be helpful.

ACKNOWLEDGMENTS

We thank Dr. Isabelle Darcy, who kindly allowed us to analyze the data for our study. We also thank John Levis, Charles Nagle, and Erin Todey and reviewers (Lucy Pickering and one anonymous reviewer) for their constructive feedback on our manuscript and the audience at PSLLT for their positive feedback on our presentation. Lastly, we acknowledge travel support from the UWM graduate school, Linguistics Department, to present our results at PSLLT.

ABOUT THE AUTHORS

Jieun Lee is a PhD student of Linguistics at the University of Wisconsin-Milwaukee (UWM). Her research interests concern L2 acquisition, phonetics & phonology, and L2 training methods.
Email: jieunlee@uwm.edu

Dong Jin Kim is a PhD student of Linguistics at UWM. His interests lie in phonetics, phonology, and prosody.
Email: kim252@uwm.edu

Dr. Hanyong Park is an associate professor at the Linguistic department at UWM. He is interested in L2 speech perception/production and effective L2 training methods.
Email: park27@uwm.edu

REFERENCES

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1-48.

Bouchard R. E., 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.

Browman, P., & Weenink, D. (2017). Praat: Doing phonetics by computer (Version: 6.0.43). Amsterdam, The Netherlands: University of Amsterdam. Retrived from http://www.praat.org/

Darcy, I., Park, H., & Yang, C. L. (2015). Individual differences in L2 acquisition of English phonology: The relation between cognitive abilities and phonological processing. *Learning and Individual Differences*, *40*, 63–72.

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

Derwing, T. M., Munro, M. J., & Thomson, R. I. (2007). A longitudinal study of ESL learners' fluency and comprehensibility development. *Applied Linguistics*, *29*(3), 359–380.

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.

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.

Douglas Fir Group. (2016). A transdisciplinary framework for SLA in a multilingual world. *The Modern Language Journal*, *100*(S1), 19–47.

Giles, H. (1970). Evaluative reactions to accents. *Educational Review*, *22*(3), 211–227.

Jaeger, B. (2017). r2glmm: computes r squared for mixed (multilevel) models. R package version 0.1.2. Retrieved from https://CRAN.R-project.org/package=r2glmm.

Jakšič, J. (2018). *Comprehensibility and pleasantness of different varieties of English as judged by Czech listeners* (Doctoral dissertation, Univerzita Karlova). Retrieved from https://dspace.cuni.cz/handle/20.500.11956/100028

Kim, D. J., Lee, J., & Park, H. (2018, October). Some candidates for evaluating fluency, accentedness, comprehensibility, and pleasantness of non-native speech. The 23rd Annual Mid-Continental Phonetics & Phonology Conference, Northwestern University, Evanston, IL.

Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, *32*(2), 145–164.

Kuznetsova A., Brockhoff, P. B., & Christensen R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1-26.

Lee, S. (2018). *Effective Planning in Real-Time Speaking Test Tasks*. Unpublished Doctoral dissertation, Michigan State University.

MacIntyre, P. D. (2007). Willingness to communicate in the second language: Understanding the decision to speak as a volitional process. *The Modern Language Journal*, *91*(4), 564-576.

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.

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.

Munro, M. J., & Derwing, T. M. (1999). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, *49*(s1), 285–310.

Munro, M. J., & Derwing, T. M. (2001). Modeling perceptions of the accentedness and comprehensibility of L2 speech the role of speaking rate. *Studies in second language* *acquisition*, *23*(4), 451-468.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133-142.

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.

R Core Team. (2017). R: A language and environment for statistical computing (Version 3.4.0). Vienna, Autria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

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.

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

Zetterholm, E., & Abelin, Å. (2017). Swedish and Somali listeners' attitudes towards L2 Swedish speech. *Scandinavian Philology*, *15*(2), 193–203.

Liu, D., & Reed, M. (2019). Asymmetrical cognitive load imposed by processing native and non-native speech. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Ames, IA, September 2018 (pp. 179-190). Ames, IA: Iowa State University.

PRESENTATION/POSTER

ASYMMETRICAL COGNITIVE LOAD IMPOSED BY PROCESSING NATIVE AND NON-NATIVE SPEECH

Di Liu, Boston University
Marnie Reed, Boston University

Intonation affects information processing and comprehension. Previous research has found that some international teaching assistants (ITAs) fail to exploit English intonation, potentially posing processing difficulties to students who are native English speakers. However, researchers have also found that non-native listeners found it easier to process sentences given by a non-native speaker with a shared language background, leading to an interlanguage speech intelligibility benefit (ISIB). Therefore, how native speaker teaching assistant (NSTA)'s and ITA's classroom speech affects the processing, comprehension, and attitudes of listeners with different language backgrounds needs to be further investigated. Using a dual-task paradigm, a comprehension questionnaire, and an attitudinal questionnaire, the present study investigates how the pronunciation and intonation of a NSTA and an ITA affect native English speakers' and Mandarin-speaking English learners' processing and comprehension of a lecture, and attitudes towards the two instructors. The present study found shared processing advantages when the listeners shared the L1 of the speaker, but overall lecture comprehension and attitude were unaffected. These findings support and extend prior research studies surveying ITAs' intonational patterns and ISIB. These findings also have implications for research on the teaching of English pronunciation to non-native instructors.

INTRODUCTION

The field of English pronunciation teaching has witnessed two profound and significant changes in the past few decades. First, scholars and teachers rejected accent-reduction oriented pronunciation teaching and embraced the intelligibility principle. Second, more attention has been given to suprasemental features such as intonation, stress, pauses, and rhythm.

Intelligibility, comprehensibility, and accentedness

Munro and Derwing (1995) conducted a study in which they asked native English speakers to listen to English sentences produced by native Mandarin speakers, transcribe the sentences, and rate each speaker's comprehensibility and accentedness. The result shows that, although it took significantly longer for the listeners to verify whether a message is correct, errors in the verification and transcription of the Mandarin-accented utterances were fairly rare (5-10%). Based on this and prior work by Abercrombie (1949), Munro and Derwing introduced three distinctly measurable concepts to second language pronunciation: Intelligibility, Comprehensibility and Accentedness. Intelligibility enhancement then became the new goal of pronunciation teaching. The following table from Derwing and Munro's (2005) study summarizes their definition and measurement of these three concepts.

Table 1

Intelligibility, comprehensibility and accentedness (Derwing & Munro, 2005, p. 385)

Intelligibility, Comprehensibility, and Accentedness

Term	Definition	Measure
Intelligibility	The extent to which a listener actually understands an utterance	Transcription task % words correct
Comprehensibility	A listener's perception of how difficult it is to understand an utterance	Scalar judgment task 1 = extremely easy to understand 9 = extremely difficult to understand
Accentedness	A listener's perception of how different a speaker's accent is from that of the L1 community	Scalar judgment task 1 = no accent 9 = extremely strong accent

While the transcription task may provide a relatively accurate measurement of intelligibility, Munro and Derwing acknowledged the complexity of intelligibility in their later studies. In particular, Munro and Derwing (2015) argued that "although the focus in L2 research is often on the characteristics of L2 speakers' productions, listeners play a crucial role in establishing the consequences of those characteristics" (p. 388). The current study, therefore, investigates how a NSTA's and a Mandarin-speaking ITA's classroom speech affects listeners who are native English speakers and Mandarin-speaking English learners. This design partially simulates university classrooms in which the instructors and the students may or may not share the same language background. The results of the present study, therefore, can help us to understand how instructors' pronunciation affects the processing, comprehension, and attitudes of students with different language backgrounds in a classroom setting.

Munro and Derwing also distinguish between local intelligibility— which refers to "how well listeners recognize relatively small units of speech, such as segments and words, outside of a larger meaningful context" (p. 381)— and global intelligibility— which entails "larger units of language that include rich contextual information" (p. 381). They argue that instructors whose goal is to enhance students' communicative competence should focus on global intelligibility, though local intelligibility proves more important for understanding the L2 acquisition process. Using a fairly long (742-word) passage delivered in the format of an online course, the present study intends to provide implications for NSTAs' and ITAs' global intelligibility within a classroom discourse context.

Suprasegmental features, prosody, and intonation

The terms "suprasegmentals" and "prosody" are commonly used interchangeably. They encompass a number of linguistic features such as intonation, pauses, stress, and rhythm. Pronunciation was formerly dominated by segmental teaching focusing on correct production of individual consonant and vowel sounds. However, research in the past thirty years clearly demonstrated that, in English, the suprasegmental features, or prosodic features, also play a crucial role in intelligibility (Anderson-Hsieh, Johnson, & Koehler, 1992, Derwing, Munro, & Wiebe, 1998). Gilbert (2014), for example, argued that "[i]n English, prosodic cues serve as navigation guides to help the listener follow the intentions of the speaker. These signals communicate emphasis and make clear the relationship between ideas (new and old information) so that listeners can readily identify these relationships and understand the speaker's meaning" (p. 123).

Within the last two decades, a growing number of researchers working on prosody have narrowed their scope and shifted their attention to the function of prosody in academic settings. Researchers have argued that prosody serves as a contextual cue and organizational device in classroom interaction and that teachers may choose different configurations of prosodic resources to help embody the type of interaction they want to accomplish and to control the direction of classroom discourse (Hahn, 2004; Hellermann, 2003; Pickering, 2004; Wennerstrom, 2001). For example, Chun (1988) states that, in classrooms, "teachers use a much wider range of communicative functions; their 'privileges' include addressing others, selecting the next speaker, choosing the topic, interrupting, asking for clarification, changing the subject, and concluding a discussion" (p.82). Wennerstrom (1998) proposed that the intonation system in English functions at the discourse level to signal relationships in information structure and mark interdependencies among constituents; she proposes a model in which intonation functions as a grammar of cohesion.

Hellermann (2003) reviewed over 25 hours of classroom IRF (initiation-response-feedback) interaction and confirmed the communicative value of prosody in a classroom. His analysis shows that teachers and students systematically use intonation in creating an effective classroom discourse. Furthermore, he found that teachers use complex prosody packaging while providing feedback to students. Skidmore and Murakami (2010) investigated an additional, important function of prosody in classroom discourse; specifically, in teacher-led dialogue, prosody might serve to signal shifts from one kind of teacher-student interaction to another. They found that prosody signals the boundaries between multiple teacher-led IRF activities, such as the passage from "thinking aloud" to an authoritative discourse used when assigning homework.

ITAs' pronunciation

Having compared NSTAs' and ITAs' intonation in classroom speech, Pickering (2001) argued that "intonation bears a high communicative load in terms of information structuring and rapport building between discourse participants" (p. 234). Pickering's (2004) analysis of NSTAs' instructional monologues reveals a hierarchy of phonologically defined units that coincide with structural boundaries at other levels of discourse; these prosodic elements contribute to the overall organization of the teaching presentations, whether the instructors recognize it or not. Her analysis of parallel ITA data, however, demonstrates that Mandarin-speaking ITAs lack effective control of English prosody at this level of structural organization. Specifically, the Mandarin speaking International teaching assistants (ITAs) "tend to create a flat, monotonic pitch structure unfamiliar to NS hearers" (Pickering, 2001, p. 249). ITAs' failure in exploiting English prosody may affect English listeners' processing of classroom speech. Gilbert (2014) noted that "emphasis that conveys the wrong meaning, or thought groups that either run together or break in inappropriate places, cause extra work for the listener who is trying to follow the speaker's meaning. If the burden becomes too great, the listener simply stops listening" (p.123).

Using a dual-task paradigm in which subjects were asked to complete a visual task while listening to a lecture, Hahn (2004) found that the ITA's correct use of sentence stress helped native speakers of English to process the information more easily. Native English speakers also recalled significantly more content and rated the speaker more favorably when the sentence stress was correctly placed. Hahn's groundbreaking study established a foundation for extending the investigation to the discourse level, and to the effects of ITAs' speech on non-native English-speaking listeners.

Interlanguage speech intelligibility benefit (ISIB)

The characteristics of ITAs' classroom speech, although negatively affecting English listeners' processing, may not pose the same challenge for listeners who are non-native speakers of English. For example, Bent and Bradlow (2003) found that non-native listeners considered speech from a high proficiency non-native speaker as intelligible as speech from a native speaker. They argued that there is an "interlanguage speech intelligibility benefit" (ISIB) between speaker and listener who share the same L1. They also found that ISIB exists even when the non-native listeners and speakers share different L1s. However, their study focuses on sentence level intelligibility; whether ISIB exists at the discourse level and in classroom discourse needs to be further investigated. Using a dual-task paradigm similar to the one used by Hahn (2004), the present study investigates how the pronunciation and intonation of a NSTA and an ITA affect processing, comprehension, and attitude of the listeners who are native English speakers and Mandarin-speaking English learners.

Research questions

The research questions are:

1. How do the pronunciation and intonation of a NSTA and an ITA affect listeners' processing and comprehension of a lecture and their attitudes toward the speakers?
2. How do listeners' L1s affect their processing and comprehension of a lecture given by an NSTA and an ITA and their attitudes toward the instructors?

METHODS

Participants

The participants are twenty-one undergraduate and graduate students recruited in a university in the US. Nine participants are native speakers of English; twelve are Mandarin-speaking English learners.

Materials

One NSTA and one native Mandarin-speaking ITA were recruited to be the speakers of this study. The two teaching assistants are both female doctoral students. The ITA is a high proficiency English speaker with a 2013 TOEFL iBT score over 100 and a U.S. master's degree. When she was asked to be a speaker of the present study, she had been living in the US for six years.

The two teaching assistants were asked to "teach an online mini lesson" with the same lecture script adapted from *Sound Concepts*, a pronunciation textbook written by Reed and Michaud (2005). A paragraph of this lecture is given below:

> Okay, today we're gonna talk about the universality of human emotions. First of all, let me say that this theory is attributed to Paul Ekman, a professor of psychology who's known as "the world's most famous face reader." Dr. Ekman's based at the University of California Medical School at San Francisco, but he's done research all over the world. Dr. Ekman says he's always been interested in emotions, ever since he was a teenager. And, being a photographer since he was twelve, he just naturally decided to look at facial expressions. In Ekman's view, it turns out there're seven basic human emotions: anger, sadness, fear, surprise, disgust, contempt, and happiness. All of these emotions have clear facial signals. There're actually 43 facial muscles that combine to reveal these emotions…

Procedure

Participants were randomized into two groups. One group listened to the lecture given by the NSTA, the other group listened to the lecture given by the ITA. Using a "dual task paradigm" to investigate the processing of the lecture by the listeners, participants were asked to accomplish two tasks simultaneously—listen to a lecture, and indicate the orientation of projected images. Based on cognitive capacity, the faster the participants complete the orientation task, the easier it is to process the lecture content.

Superlab, a presentation software, was used to deliver the lectures. While the participants were listening to the lecture, they saw images that appeared every five seconds on the screen. They were asked to press either of two keys to indicate for each image whether it was upright (press "j") or inverted (press "f") as quickly as possible. The participants were instructed that their primary task was to comprehend and remember the content of the lecture, and their secondary task was to press the correct key as quickly as they can.

After listening to the lecture, participants were asked to complete seven open-ended comprehension questions. The questions tested participants' comprehension of the main idea and details in the lecture (e.g., Why did Ekman go to Papua New Guinea?)

At the end of the study visit, participates were asked to complete a speaker evaluation questionnaire adapted from Hahn (2004). The participants were asked to rate the instructors based on a 5-point Likert scale. Two sample questions are provided below:

Q2: How would you characterize the instructor's ability to explain?

Excellent 5 4 3 2 1 Very poor

Q5: It was easy to hear and understand the instructor.

Almost always 5 4 3 2 1 Almost never

Data analysis

Processing time (in milliseconds) was analyzed and compared across groups using the R (R Core Team, 2012) and the lme4 package (Bates, Maechler, & Bolker, 2012). Four groups were derived from combinations of the teaching assistants' and the participants' language backgrounds (i.e., native English speakers (English S) listening to NSTA, native English speakers (English S) listening to ITA, native Mandarin speakers (Mandarin S) listening to NSTA, and native Mandarin speakers (Mandarin S) listening to ITA). These four combinations were compared using a linear mixed-effects model. The combinations of speaker and listener language background were entered as the fixed effect. Participants' ID and trial number were entered as random effects.

Each question in the comprehension questionnaire was assigned a score of three or four. The total score of the comprehension questionnaire is 27. For example, in the question, "According to Ekman, how many universal emotions do human beings have and what are they?", participants got one point for every two emotions they wrote down. If they wrote down all seven emotions, they got four points. A one-way between-subject ANOVA was conducted to compare the comprehension questionnaire scores among different combinations of speakers' and listeners' language background.

The attitudinal questionnaire used 5-point Likert scales to obtain speaker ratings. The total scores were analyzed based on a one-way between-subjects ANOVA.

RESULTS

Processing

Compared to the native English listeners listening to the NSTA, the native English listeners listening to the ITA had significantly longer reaction time, the native Mandarin listeners listening to the NSTA also had significantly longer reaction time. Compared to the English listeners listening to the ITA, the native Mandarin listeners listening to the ITA had significant shorter reaction time. However, there is no significant difference between the reaction time of the English listeners listening to the NSTA and the Mandarin listeners listening to the ITA (see Table 2).

Table 2

Comparison of processing time across different combination groups.

Contrast	Estimate	SE	df	t-value	p-value
English-NSTA vs. English-ITA	-91.62	18.05	1320.31	-5.08	<0.0001***
English-NSTA vs. Mandarin-NSTA	-62.81	16.97	1290.38	-3.70	0.0013*
English-NSTA vs. Mandarin-ITA	-25.78	16.42	1328.59	-1.57	0.39
English-ITA vs. Mandarin-NSTA	28.81	17.69	1325.09	1.63	0.36
English-ITA vs. Mandarin-ITA	65.83	17.14	1290.35	3.84	0.0007**
Mandarin-NSTA vs. Mandarin-ITA	37.02	15.32	1329.06	2.42	0.07

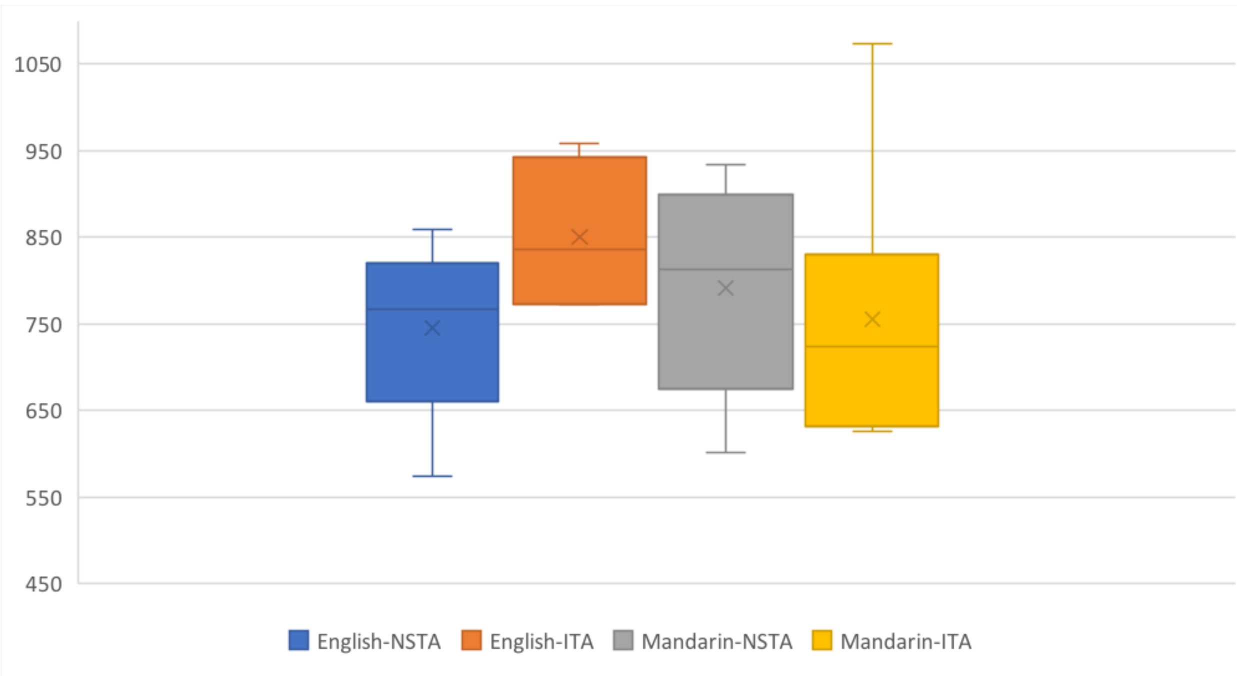


Figure 1. Processing time in the dual-task paradigm.

Comprehension

A one-way between-subjects ANOVA was conducted to compare the effect of listeners' language background on comprehension scores. The results show there was not a significant effect of listeners' language background on comprehension scores [F (3, 17) = 0.36, p = 0.78]. The post hoc analysis revealed that the statistical power for this analysis was 0.84. This result suggests that despite different processing difficulties, listeners' comprehension of the lecture content might not be significantly affected by the speakers' language background.

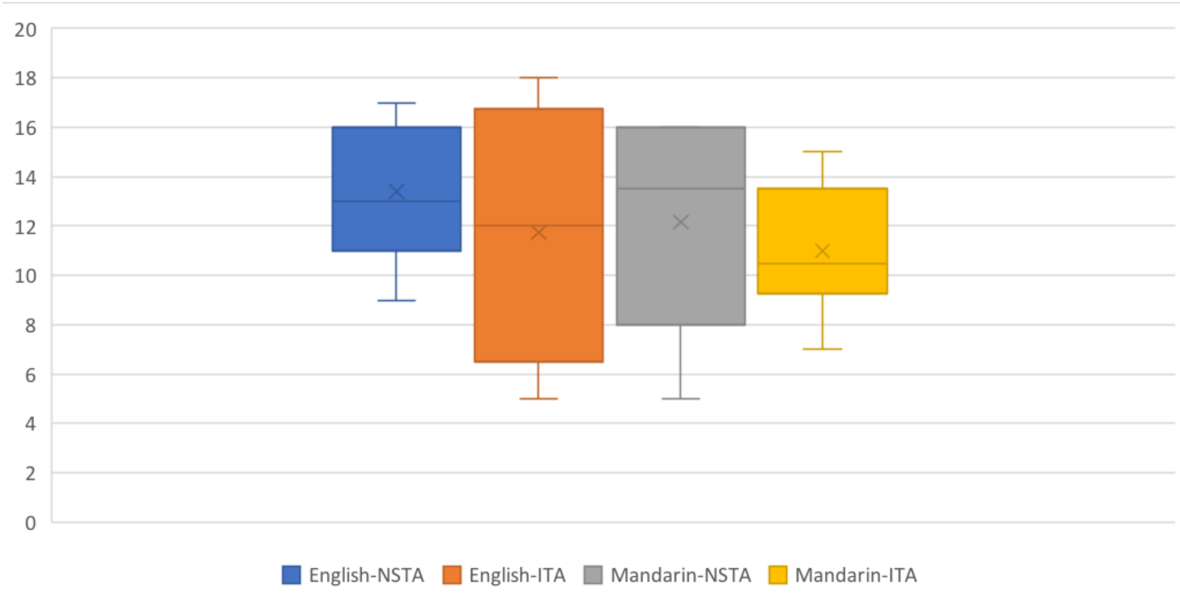


Figure 2. Comprehension scores of the four combination groups.

Attitude

A one-way between-subjects ANOVA was conducted to compare the effect of listeners' language background on attitude scores. The results show there was not a significant effect of listeners' language background on listeners' attitude towards the speakers [F (3, 17) = 1.38, p = 0.28]. The post hoc analysis revealed that the statistical power for this analysis was 0.99.

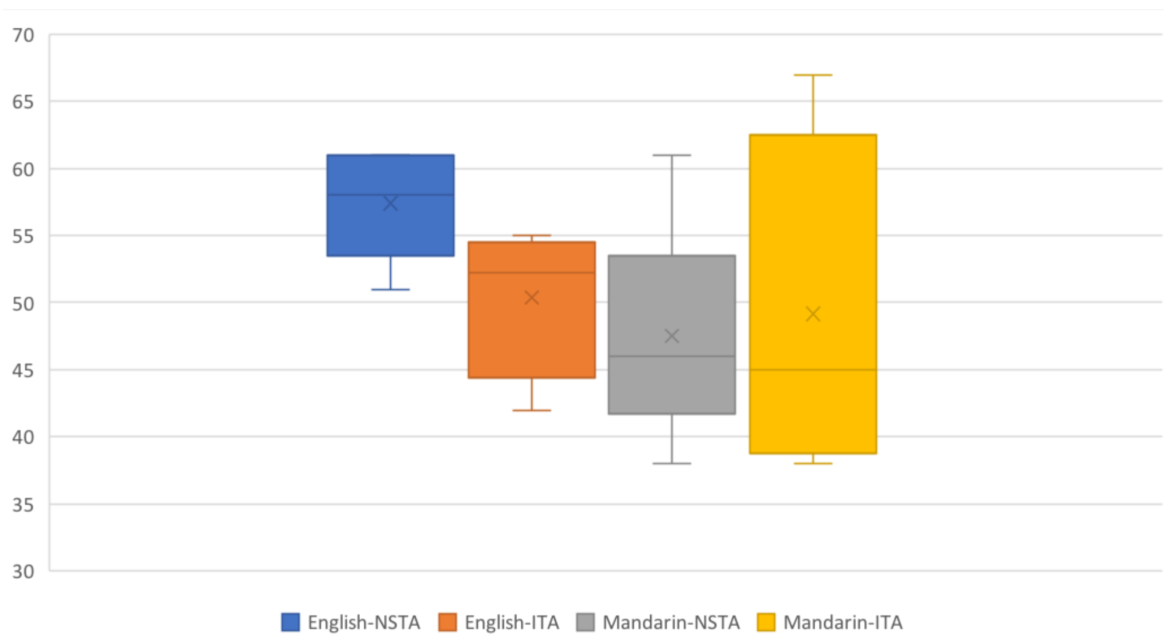


Figure 3. Attitudinal questionnaire scores of the four combination groups.

DISCUSSION

Using a dual-task paradigm, a comprehension questionnaire, and an attitudinal questionnaire, this study investigated whether native speaker teaching assistant (NSTA)'s and ITA's speech affects the processing, comprehension, and attitudes of listeners with different language backgrounds. The results show that the NSTA's and ITA's speech poses asymmetrical processing difficulties to listeners with different L1s. Specifically, it is easier for the native listeners to process the lecture given by the NSTA and it is easier for the non-native listeners than it is for the native listeners to process the lecture given by the ITA. These results confirmed prior researchers' findings on ITAs' pronunciation deficiency (Pickering, 2001) and the Interlanguage Speech Intelligibility Benefit (Bent & Bradlow, 2003) and extended them to the discourse level using a classroom lecture. However, the "mismatch interlanguage speech intelligibility benefit", which suggests that it is easier for non-native listeners to process the speech of a non-native speaker even when the listener and the speaker have different L1s, was not investigated in this study. It remains unsolved and should be investigated in future studies. Despite the difference in processing difficulties, there are no statistically significant differences in listeners' comprehension of the lecture or their attitudes towards the speakers. These results suggest that the processing difficulties may not necessarily lead to inferior learning outcomes or negative attitudes towards the speaker.

Some scholars attribute non-native speakers' deficiency in exploiting English pronunciation to the differences in the intonational system of English and the learners' L1. For example, Clennell (1997) argues that "the discourse/pragmatic functions of English prosody appear to be specific to the English language and are unfamiliar to most learners of English" (p.117). However, recent studies challenge this claim. For instance, scholars have found many similarities between English and Mandarin prosody, from physical correlations to pragmatic functions (Chen & Gussenhoven, 2008; Ip and Cutler, 2016; Liu, 2017; Ouyang & Kaiser, 2015). As Ip and Cutler (2016) concluded, "Information structure is a linguistic universal… all speakers have the option to convey this structure in the way they speak, and they may use prosody to do it" (p. 330). But if English and the learners' L1 share similar prosodic features and functions, why is it that non-native speakers do not transfer these functions to English? Multiple factors could lead to the lack of positive transfer.

It is possible that the non-native speakers are unaware of the functions and importance of prosody, both in their own language and in English. As Gilbert said, the importance of prosody is not "self-evident." In this case, raising learners' awareness of the importance and the pragmatic functions of English prosody might improve non-native speakers' use of English prosody, which could make it easier for native listeners to process the ITAs' speech.

It is also possible that even though both English speakers and Mandarin speakers use prosody in a similar manner and to a similar extent, English listeners rely more heavily on prosodic cues. This may explain why native listeners found it particularly challenging to process the ITA speech whereas non-native listeners, relying less heavily on prosodic cues, did not find it more challenging to process ITA speech.

Familiarity is another issue that could lead to the asymmetrical cognitive load imposed by the pronunciation and intonation difference of the NSTA's and ITA's speech. It has been found by prior researchers (Gass & Varonis, 1984) that listeners' familiarity with the speakers' accent is a key variable in determining the intelligibility and comprehensibility of the speaker. Specifically, a non-native speaker's speech may be more intelligible to a listener who is familiar with the non-native speaker's accent. It is possible that the participants who are native speakers of English are less familiar with the Mandarin-speaking ITA's accent whereas the participants who are non-native speakers of English are familiar with both the Mandarin-speaking ITA's accent and the NSTA's accent. In other words, native listeners may have found it harder to process the ITA's speech because they are less familiar with Mandarin-accented English pronunciation.

ABOUT THE AUTHORS

Di Liu is a part-time lecturer at Boston University. His research focuses on cross-linguistic comparison of prosody, cognitive influence of L1 and L2 prosody, and technology-enhanced pronunciation teaching.

Marnie Reed is a professor of education at Boston University. Her research focuses on second language phonology. Her current areas of interest include the role of metacognition in cross-linguistic awareness of the pragmatic functions of English prosody.

REFERENCES

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

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,* 529–555.

Bates, D., Maechler, M., & Bolker, B. (2012). lme4: Linear mixed-effects models using S4 classes. CRAN. R project.

Bent, T., & Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. *The Journal of the Acoustical Society of America*, *114*(3), 1600-1610.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5*(9-10), 341-345.

Chen, Y., & Gussenhoven, C. (2008). Emphasis and tonal implementation in Standard Chinese. *Journal of Phonetics*, *36*(4), 724-746.

Chun, D. (1988). Teaching intonation as part of communicative competence: Suggestions for the Classroom. *Die Unterrichtspraxis / Teaching German, 21*(1), 81-88.

Clennell, C. (1997). Raising the pedagogic status of discourse intonation teaching. *ELT Journal, 51*(2), 117-25.

Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly, 39*(3), 379-397.

Derwing, T. M., Munro, M. J., & Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language learning, 48*(3), 393-410.

Gass, S., & Varonis, E. M. (1984). The effect of familiarity on the comprehensibility of nonnative speech. *Language learning*, *34*(1), 65-87.

Gilbert, J. (2014). Myth 4: Intonation is hard to teach. In L. Grant (Ed.), *Pronunciation myths: Applying second language research to classroom teaching* (pp.107-136). Ann Arbor, MI: University of Michigan Press.

Liu, D. (2017). Mandarin speaker's intonational emphasis in English and Mandarin lectures. In M. O'Brien & J. Levis (Eds). *Proceedings of the 8th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Calgary, AB, August 2016 (pp. 85-96). Ames, IA: Iowa State University.

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.

Munro, M. J., Derwing, T. M. (2015). Intelligibility in research and practice: Teaching priorities. In M. Reed & J. M. Levis (Eds.), The Handbook of English pronunciation (pp. 377–396). West Sussex, UK: John Wiley & Sons, Inc.

Hahn, L. D. (2004). Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly, 38*(2), 201-223.

Hellermann, J. (2003). The interactive work of prosody in the IRF exchange: Teacher repetition in feedback moves. *Language in Society, 32*(1), 79-104.

Ip, M. H. K., & Cutler, A. (2016). Cross-language data on five types of prosodic focus. *Proceedings of the Speech Prosody Conference 2016*, 330-334.

Ouyang, I. C. & Kaiser, E. (2015). Prosody and information structure in a tone language: An investigation of Mandarin Chinese. *Language, Cognition and Neuroscience*, *30*(1-2): 57-72.

Pickering, L. (2001). The role of tone choice in improving ITA communication in the classroom. *TESOL Quarterly, 35*(2), 233-255.

Pickering, L. (2004). The structure and function of intonational paragraphs in native and nonnative speaker instructional discourse. *English for Specific Purposes, 23*(1), 19-43.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Reed, M., & Michaud, C. (2005). *Sound concepts: An integrated pronunciation course*. New York: McGraw-Hill.

Skidmore, D., & Murakami, K. (2010). How prosody marks shifts in footing in classroom discourse. *International Journal of Educational Research, 49*(2-3), 69-77.

Wennerstrom, A. (1998). Intonation as cohesion in academic discourse: A study of Chinese speakers of English. *Studies in Second Language Acquisition, 20*(1), 1-25.

Wennerstrom, A. (2001). *The music of everyday speech: Prosody and discourse analysis*. Oxford University Press.

McCrocklin, S., Humaidan, A., & Edalatishams, E. (2019). ASR dictation program accuracy: Have current programs improved? In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Ames, IA, September 2018 (pp. 191-200). Ames, IA: Iowa State University.

PRESENTATION/POSTER

ASR DICTATION PROGRAM ACCURACY: HAVE CURRENT PROGRAMS IMPROVED?

Shannon McCrocklin, Southern Illinois University
Abdulsamad Humaidan, Southern Illinois University
Idée Edalatishams, Iowa State University

Automatic Speech Recognition (ASR) dictation programs have the potential to help language learners get feedback on their pronunciation by providing a written transcript of recognized speech. Early research into dictation programs showed low rates of recognition for non-native speech that prevented usable feedback (Coniam, 1999; Derwing, Munro, & Carbonaro, 2000), but updated research revisiting the accuracy of dictation transcripts for non-native speech is needed. This study investigates current accuracy rates for two programs, *Windows Speech Recognition* (*WSR*) and *Google Voice Typing (Google)*. Participants (10 native English speakers and 20 advanced non-native speakers) read 60 sentences and responded to two open-ended questions. Transcripts were analyzed for accuracy and *t*-tests were used to make comparisons between programs. Major findings include: 1) *Google* displayed a tendency to turn off in the middle of transcription, which affected rates of attempted words; 2) when comparing the accuracy for native versus non-native speech, both programs had higher levels of accuracy for native speech; and 3) when comparing programs for the same speaker, *Google* outperformed *WSR* for both speaker groups on both tasks. Comparing the results to Derwing et al. (2000), *Google* seems to offer substantial increases in accuracy for non-native speakers.

INTRODUCTION

Automatic Speech Recognition (ASR) is a “machine-based process of decoding and transcribing oral speech” (Levis & Suvorov, 2012, p. 1) that is built into numerous technologies such as automated call centers and dictation programs. ASR technology is also common in language learning software, such as *Rosetta Stone*. ASR has been an interest in the field of pronunciation training since the 1990s partially thanks to the reemerging interest in developing students’ spoken language skills (Cucchiarini & Strik, 2018).

Much of the early interest in ASR focused on dictation programs. Dictation programs were developed for native speakers of a given language and are built into both Windows and Mac operating systems as part of their accessibility services. Dictation programs use ASR to interpret what the user has said and provide the spoken utterance in written form. Early tests of the potential of dictation programs for pronunciation practice in a second language highlighted concerns about both the accuracy of the programs for non-native speakers as well as the usability of the feedback (Coniam, 1999; Derwing et al. 2000). Derwing et al. (2000) asked 30 participants (10 native speakers of English, 10 Spanish L1 speakers, and 10 Chinese L1 speakers) to read 60 sentences to *Dragon Naturally Speaking*, a dictation program utilizing ASR. The researchers then assessed the transcription against known intended sentences and against human listeners. They found that while human listeners were able to understand 95% of the words produced by non-native speakers (and almost 100% of the words produced by native speakers), *Dragon's* accuracy rate was much lower, 71-72% for non-native speakers and 90% for native speakers. Researchers concluded that, given the high levels of inaccurate transcription, use of dictation transcripts would lead to unreliable feedback for second language learners.

In subsequent years, the field instead turned its attention to Computer Assisted Pronunciation Training (CAPT). CAPT programs are developed specifically for non-native speakers of a given language. In a CAPT program, utterances are controlled by having the participant read a written text or respond to a limited range prompt (Hincks, 2015). The program then compares the ASR recognition to the intended response in order to formulate a score or feedback for the student. Within CAPT programs, great strides have been made to improve the accuracy of the evaluation of speech by performing acoustic analysis (Truong, Neri, de Wet, Cucchiarini, & Strik, 2005), incorporating data from non-native speakers (Bouselmi, Fohr, & Illina, 2012; Moustroufas & Digalakis, 2007), examining changes in pronunciation when words are used as part of a larger discourse (Saraçlar, 2000), and hierarchically ranking pronunciation issues based on saliency for more useful feedback (Tepperman, 2009). More importantly, research has shown that CAPT programs can facilitate learning for diverse populations of learners (e.g. children and adults as well as different language backgrounds) (Hincks, 2003; Neri, Cucchiarini, & Strik, 2006; Neri, Mich, Gerosa, & Giuliani, 2008). CAPT programs are, however, limited in their flexibility. Students must follow along prescribed plans of study designed into the CAPT program and must only practice the pre-programmed utterances.

Dictation programs, on the other hand, allow learners to work on whatever topic or language item they wish to. Students could practice words that they struggled with, speak freely to the program on new topics, practice presentations for class, emulate famous speeches, or even read poetry to a dictation program. In recent years, researchers have redeveloped interest in dictation programs for pronunciation practice. Recent research has shown that dictation practice can facilitate student improvement. Research examining practice with dictation programs found that students can improve their production of segmentals using dictation practice with ASR equally well (McCrocklin, 2019) and perhaps even better than when experiencing face-to-face instruction (Liakin, Cardoso, & Liakina, 2014). McCrocklin (2019) focused on a variety of English segmentals (both consonants and vowels) following practice with *Windows Speech Recognition* (*WSR*), while Liakin, Cardoso, & Liaking (2014) focused on the French vowel /y/ following practice with *Dragon Dictation*, a mobile dictation application. However, benefits of ASR practice may extend beyond student improvement. Students reported feeling more empowered in their pronunciation practice when exposed to ASR-based dictation practice (McCrocklin, 2016) and, after using *Google Web Speech (Google)* with pronunciation students, Wallace (2016) argued that ASR dictation practice is useful for getting students to notice pronunciation errors. Wallace described having students dictate while also recording themselves speaking. Students then worked to correct the dictated transcript, using the recording to check what was originally said and to allow for analysis of pronunciation errors. Finally, Mroz (2018) found that students felt that ASR provided a measure of intelligibility useful for understanding how native speakers may perceive their speech.

Despite the recent resurgence of interest, it is unclear to what degree dictation programs have improved in accuracy over the years. McCrocklin (2016) noted that students were still quite frustrated by the number of mistranscribed words provided by *WSR*. To begin answering this question, Edalatishams (2017) compared *Dragon Naturally Speaking* and *Mac Dictation* with 12 sentences (total 58 content words) read aloud by 12 NNSs, finding that Mac Dictation had an average accuracy rate of 77% while Dragon Naturally Speaking had an average accuracy rate of 72%, which was much lower than findings for human listeners (89-98%). These results suggest that perhaps programs have not substantially improved. Without substantial improvement, it is unlikely that programs have moved closer to the goal of providing indications of human intelligibility levels. However, more data is needed and more programs should be tested for accuracy analyses with non-native speakers, ideally using larger data samples.

THE PRESENT STUDY

The present study examines ASR dictation with specific attention to three dimensions: speakers, dictation program, and speech task. In particular, the preliminary analysis seeks to understand whether changes to speakers, program used, and speech task makes a significant difference to the accuracy of the dictation. Native and non-native speakers of English used *Google* and *WSR* for transcribing read-aloud sentences and free responses.

Participants

The study included 10 native speakers of American English and 20 non-native speakers whose first languages were Spanish, Chinese, Arabic, French, and Ewe. For this preliminary study, the sample was primarily one of convenience, but efforts were made to gather participants from a variety of language backgrounds. The non-native speakers were advanced English language learners who entered the university with an average TOEFL score of 89.3. The average age of participants, both native and non-native participants was 25.4. The majority of participants were female (60%) in both groups. See Table 1 for more information about each group.

Table 1

Participant details by language background

	Native Speakers	Non-native Speakers
Number of participants	n=10	n=20
Average Age	24	26
Gender	Female n=6	Female n=12
	Male n=4	Male n=8
Native Language	English n= 10	Spanish n=7
		Chinese n=6
		Arabic n=5
		French n=1
		Ewe n=1

Procedures

Participants were provided with information about the study and signed an informed consent form. They were then asked for demographic information through a paper-based questionnaire. For the next stage, participants were recorded as they dictated 60 sentences to two different programs simultaneously, *WSR* and *Google Voice Typing (Google)* in *Drive*, each running on a different computer. To ensure a stable internet connection, which was important for facilitating recognition in *Google*, the computers were hard connected to campus-provided ethernet. Both computers used Logitech USB microphones which were positioned in front of the participant. Participants were also recorded on one computer using *Audacity*. After completing the 60 sentences, participants also responded to two open-ended questions and participated in correcting a copy of the transcript. More information about the dictation and open-response tasks is included in the following sections.

Dictations. The participants read aloud sixty sentences to *WSR* and *Google* on the computers, repeating each sentence twice. The second reading provided an opportunity, if needed, for participants to correct a mistake in their reading of the sentence. This step was considered useful for future analyses, although both sentences were included in this pilot and preliminary analysis. The sentences were true/false sentences used successfully in Derwing et al. (2000), adapted for use in this study with the permission of Derwing and Munro. The sentences feature a variety of topics/vocabulary as well as a variety of sounds. The average sentence length was 6.13 words per sentence. For example, a true sentence included was “You can see animals at the zoo.”

Open-ended responses. The participants also responded to two open ended questions. The dictation programs were turned off during the provision of directions in this stage as well as during the introduction of each open-ended question. Participants were asked to describe how they would plan a surprise birthday party for a friend and to describe their favorite things to do on the weekend. Participants were provided guidance to provide either about 30 seconds of speech or 4-5 sentences in response. After participants had responded to each question, *Audacity* was stopped and both transcribers were turned off. The researcher then made a copy of the transcript provided by one of the dictation programs for both of the open-ended questions. The researcher worked with participants to correct the copied transcripts to provide an accurate transcription of what had been said (to allow for comparisons with the dictated version). During this final step, participants listened to their responses recorded in *Audacity* to remember each utterance and identify mistakes in the transcript.

Analysis

Analysis included counting the number of words attempted by the program as well as the accuracy of each sentence. Using the accurate list of sentences provided for reading and the accurate, corrected copy of the transcript for the open-ended responses, the ASR-dictated transcripts were scored for the number of accurate words successfully transcribed from the correct version. For this preliminary analysis, the first 10 read sentences, using an average of the first and second attempt, and the first 5 sentences of the open-ended responses were analyzed. Because *Google* had a tendency to turn off, we calculated average number of attempted words as a percentage within each sentence as well as the accuracy of the transcription within the attempted words. For example, if a sentence had 10 words, but *Google* turned off after five, it would have an attempt rate of 50%.

If out of those five, *Google* got four correct, it would have an accuracy rate (within the attempted words) of 80% or 4/5. To compare the accuracy for native speech versus non-native speech on each program (*WSR* and *Google*) for each task (read speech and open-ended response) an independent samples *t*-test was used. Paired samples *t*-tests were used to compare *WSR* and *Google* for 1) each speaker group (native and non-native) and 2) each task (controlled read speech or free open-ended response).

RESULTS

As mentioned in the analysis section, *Google* had the tendency to turn off frequently during each participant's dictation work. It is not transparent what triggered this issue, but we noticed that hitting enter after each utterance to begin each new sentence on a new line exacerbated the problem. Although we discontinued this practice, *Google* continued to turn off at unpredictable intervals during recoding. Each time it turned it off, we worked to turn it back on at the beginning of a new sentence. The first noticeable finding, then, was that when you include simply a count of accurate transcriptions (thereby counting the non-attempted words against *Google*'s number of accurate transcriptions) *Google* and *WSR* occasionally showed similar numbers of accurate words on certain measures. For example, when examining non-native speakers' read sentences, the accuracy of the words transcribed (when shown as a percentage of all possible read words) was 73.02% for *Google* and 72.1% for *WSR*. In this example, *Google* only attempted 81.04% of the words because of its tendency to shut off, while *WSR* attempted 96.74%. *Google* had a much smaller rate of inaccurate transcriptions at 7.98%, however, while *WSR* showed 24.64% inaccurate transcriptions (see Figure 1).

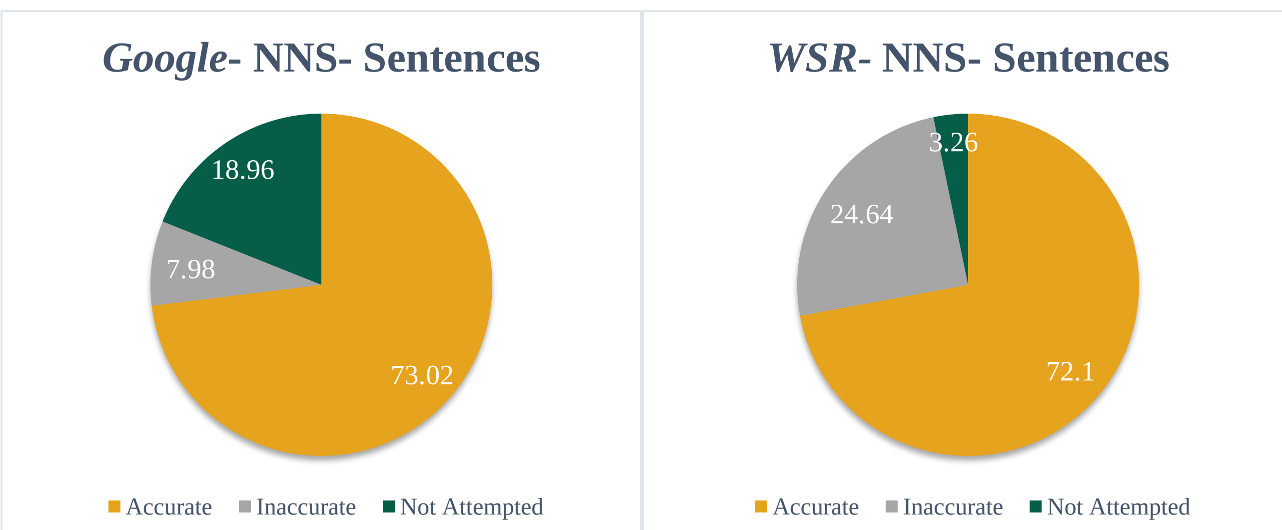


Figure 1. Percentage of accurate, inaccurate, and not attempted words for sentences read by non-native speakers by *Google* (left) and *WSR* (right).

Thus, it was important moving forward, to focus on providing the percent of attempted words per sentence as well as the percent of accurate words as a count within the attempted words. Table 2 shows the percentage of attempted words and accuracy among the attempted words for each program on each task by each speaker group (native and non-native).

Table 2

Mean percentage of attempted and accurate words (within attempted words) by program, task, and speaker group

	Sentences						Free Speech					
	Google			*WSR*			*Google*			*WSR*		
	Attempted	Accuracy Mean	Accuracy SD	Attempted	Accuracy Mean	Accuracy SD	Attempted	Accuracy Mean	Accuracy SD	Attempted	Accuracy Mean	Accuracy SD
Native	75.74	91.95	11.25	96.72	86.81	5.14	100.00	98.00	2.58	94.50	59.70	23.41
Non-native	81.04	88.61	10.43	96.74	74.44	13.42	98.82	93.47	8.30	97.06	53.50	32.11

An independent samples *t*-test was used to compare the accuracy rates for native and non-native speakers on a single program and task. The results showed statistically significant differences between the accuracy rates for *WSR* on the sentences task ($t(28)= 2.79$, $p=.002$) and for *Google* on the free speech task ($t(25)=1.67$, $p<.001$) when comparing native and non-native speakers. Specifically, both *WSR* and *Google* had higher mean accuracy rates for native speakers than for non-native speakers (86.81% vs. 74.4% for *WSR* on sentences and 98% vs. 93.47% for *Google* on free speech), which is in line with previous research from Derwing et al. (2000). In contrast, there were no statistically significant differences when comparing the accuracy of dictation for native versus non-native speakers for *Google* on sentences ($t(28)= .807$, $p=.779$) and *WSR* on free speech ($t(25)= .531$, $p=.470$). This is a surprising finding given that Derwing et al. (2000) previously found significant differences between accuracy rates for native and non-native speakers.

A paired samples *t*-test was used to compare accuracy rates for *Google* and *WSR* for the same speaker population on the two tasks. In all cases, *Google* outperformed *WSR*. The results showed statistically significant differences in three out of the four pairings, non-native speakers on the sentences task ($t(19)=5.42$, $p<.001$) and both native and non-native speakers on the free speech task ($t(9)=5.19$, $p=.001$ and $t(15)=5.14$, $p<.001$ respectively). The differences between *Google* and *WSR* on sentences for native speakers was not statistically significant ($t(9)=1.26$, $p=.238$).

The results further identified an interesting trend: While *Google* became more accurate as speakers switched from the sentence reading task to the free speech, *WSR* displayed an opposite trend. Using a paired *t*-test to compare results from the same speaker on the controlled sentence reading and the free speech task, both trends were statistically significant. For *Google*, in which the system was more accurate in the free speech task, the *p* value was .047 ($t(26)=2.09$), while in *WSR*, in which the system was more accurate in the controlled sentence reading task, the *p* value was less than .001 ($t(26)=-4.78$). Notably, *WSR*'s accuracy on free speech was barely over 50% for non-native speakers. This relationship is illustrated in Figure 2. It is not clear, however, what may have led to such disparities. It is likely that differences in the underlying programming or gaps in the training (voices/styles used to train and check the program during development) have created differential responses to differing speech features.

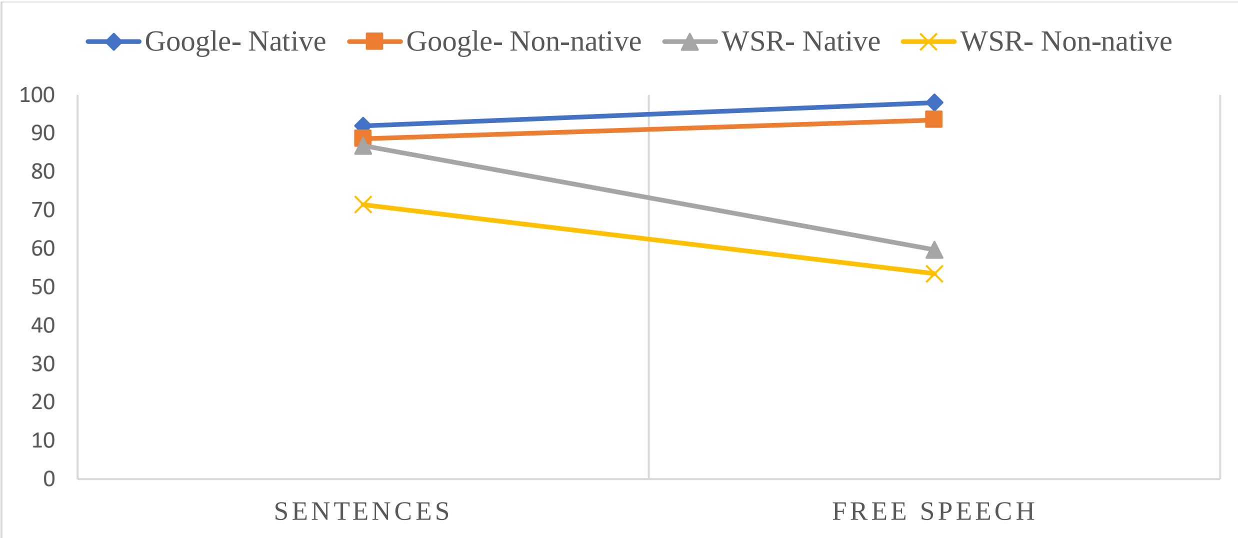


Figure 2. Accuracy of attempted words for each program and group of speakers by task type.

DISCUSSION AND CONCLUSION

The results of this study highlighted several important findings. First, attempted words was an important measure to track. *Google* had the frustrating tendency to turn off which was particularly problematic in the read sentences section, despite changes to protocol to limit the number of stoppages (i.e. forgoing the use of enter to start each sentence on a new line). Notably, although *WSR* outperformed *Google* in attempted words for the sentences task, *Google* outperformed *WSR* in the free speech task. It is not clear what behaviors may have contributed to this difference. Perhaps, because participants were speaking in shorter stretches (usually 4-5 sentences per question) with the dictation programs turned off and restarted for each, *Google* simply had less time to stop working.

Second, despite *Google*'s weakness in turning off, it had much greater accuracy within the attempted words. *Google*'s accuracy for non-native speech ranged from 88.61% (sentences) to 93.47% (free speech), while *WSR*'s accuracy for non-native speakers ranged from 53.50% (free speech) to 74.44% (sentences). *Google*'s outperformance of *WSR* was statistically significant in three out of the four pairings of speaker and task. Comparing our findings to the accuracy rates reported in Derwing et al. (2000), *Google* seems to offer substantial improvement from *Dragon Naturally Speaking* 20 years ago, which offered accuracy rates for non-native speech around 71-72%. *WSR* does not show such improvements.

Further, while the current study did not include comparisons with human raters, *Google* may be getting close to the levels of accuracy of native listeners. While Derwing et al. (2000) found that human listeners were able to transcribe 95-97% percent of non-native speech accurately, Edalatishams (2017) found a range from 89-98% for non-native speech. *Google*'s transcription levels of 88.61% (sentences) to 93.47% (free speech) suggests that *Google* may now rival human listeners particularly for free speech. Further testing is needed and planned for the audio samples in this particular study, however, to make true comparisons to human listeners. Additionally, analysis of speech samples by pronunciation experts in order to examine whether mis-transcriptions were linked to pronunciation errors will help to determine the usefulness of *Google* for use in second language pronunciation practice. However, an initial examination does show potential for *Google* to be a useful tool. Two examples are illustrated in Table 3 below.

Table 3

Original sentence, phonetic transcription of utterance, and transcription from Google for 2 participants' utterances

	Original Sentence	Phonetic Transcription of Participant Utterance	Sentence Transcription provided by *Google*
Male- Arabic L1 (PSpr18-3)	Most desks are made from spaghetti.	[most dɪsks ɑɹ med fɹəm ˈspaˌgɛti]	Most disks are made from spaghetti.
Female- Chinese L1 (PFa17-8)	You can see animals at the zoo.	[ju kæn si ˈæˌniməs æt di zu]	You can see animals HD 2.

In this example, although both speakers display multiple pronunciation features that could be labelled as errors, such as the full vowel and stress in the first syllable of "spaghetti" or the mis-stressing and lacking [l] in "animals," *Google* had trouble with "disks" because of the heightened vowel and minimal pair "desks-disks" for the male Arabic speaker. *Google* also had trouble with the full vowels and stressing of "at the" in the female Chinese speaker's utterance which should have been de-stressed as function words. This may additionally indicate that, as *Google*'s ASR has improved, it is less sensitive to accent and more likely to make errors in transcription in places where intelligibility may be negatively impacted for native speakers (for example, in instances of minimal pairs). Having a program that replicated intelligibility for human listeners, as Mroz (2018) has suggested is becoming possible, would be a huge move forward for ASR dictation, making dictation practice more useful for second language learners. Further analysis and testing is needed.

ACKNOWLEDGEMENTS

We would like to thank Tracy Derwing and Murray Munro for providing the sentences originally used in Derwing et al. (2000) and for allowing their adaptation into this study.

ABOUT THE AUTHORS

Shannon McCrocklin (shannon.mccrocklin@siu.edu) is an Assistant Professor at Southern Illinois University in the Department of Linguistics. Her primary research area focuses on second language pronunciation teaching methods and, specifically, ways of incorporating technology into second language learning.

Contact Information:
1000 Faner Dr., Rm 3228
Southern Illinois University
Carbondale, IL 62901
1-618-453-3428

Abdulsamad Humaidan (humaidanabdulsamad@siu.edu) is a doctoral candidate at Southern Illinois University pursuing his PhD in Education focusing on Curriculum and Instruction and TESOL. His research interests include ESL and EFL assessment, teacher education, and language teaching and learning with technology.

Idée Edalatishams (edalati@iastate.edu) is a graduate student at Iowa State University pursuing her PhD in Applied Linguistics and Technology. Her research is in the fields of sociolinguistics and discourse analysis using corpus methodology with a primary focus on discourse intonation and non-native speakers' identity.

REFERENCES

Bouselmi, G., Fohr, D., & Illina, I. (2012). Multilingual recognition of non-native speech using acoustic model transformation and pronunciation modeling. *International Journal of Speech Technology*, *15*, 203-213.

Coniam, D. (1999). Voice recognition software accuracy with second language speakers of English. *System*, *27*, 49-64.

Cucchiarini, C. & Strik, H. (2018). Automatic Speech Recognition for second language pronunciation training. In O. Kang, R. I. Thomson, & J. M. Murphy (Eds.), *The Routledge handbook of contemporary English pronunciation* (pp.556-569). New York, NY: Routledge.

Derwing, T., Munro, M., & Carbonaro, M. (2000). Does popular speech recognition software work with ESL speech? *TESOL Quarterly*, *34*(3), 592-603.

Edalatishams, I. (Sept, 2017). *Non-native speech and recognition accuracy of two ASR applications: Dragon and Dictation*. Paper presented at the meeting of Pronunciation in Second Language Learning and Teaching, Salt Lake City, UT.

Hincks, R. (2015). Technology and leaning pronunciation. In M. Reed & J. Levis (Eds.), *The handbook of English pronunciation* (pp. 505-519). Malden, MA: John Wiley & Sons, Inc.

Hincks, R. (2003). Speech technologies for pronunciation feedback and evaluation. *ReCALL*. *15*(1), 3-20.

Liakin, D., Cardoso, W., & Liakina, N. (2014). Learning L2 pronunciation with a mobile speech recognizer: French /y/. *CALICO Journal*, *32*(1), 1-25.

Levis, J. & Suvorov, R. (2012). Automated speech recognition. In C. Chapelle (Ed.), *The encyclopedia of applied linguistics*. Retrieved from http://onlinelibrary.wiley.com/

McCrocklin, S. (2016). Pronunciation learner autonomy: The potential of Automatic Speech Recognition. *System*, *57*, 25-42.

McCrocklin, S. (2019). ASR-based dictation practice for second language pronunciation improvement. *Journal of Second Language Pronunciation*, 5(1), 98-118.

Moustroufas, N. & Digalakis, V. (2007). Automatic pronunciation evaluation of foreign speakers using unknown text. *Computer Speech and Language*, *21*, 219-230.

Mroz, A. (2018). Seeing how people hear you: French learners experiencing intelligibility through automatic speech recognition. *Foreign Language Annals*, *1*(21), 1-21.

Neri, A., Cucchiarini, C., & Strik H. (2006). ASR-based corrective feedback on pronunciation: does it really work? *Proceedings of the ISCA Interspeech 2006*, Pittsburgh, PA, 1982–1985.

Neri, A., Mich, O., Gerosa, M., & Giuliani, D. (2008). The effectiveness of computer assisted pronunciation training for foreign language learning by children. *Computer Assisted Language Learning*, *21*(5), 393-408.

Saraçlar, M. (2000). *Pronunciation modeling for conversational speech recognition* (Doctoral dissertation). Ann Arbor, MI: UMI Dissertation Services.

Tepperman, J. (2009). *Hierarchical methods in automatic pronunciation evaluation.* (Doctoral dissertation). Ann Arbor, MI: UMI Dissertation Services.

Truong, K., Neri, A., de Wet, F., Cucchiarini, C, & Strik, H. (2005). Automatic detection of frequent pronunciation errors made by L2-learners. *Proceedings from InterSpeech 2005 (IS2005),* Lisbon, Portugal, 1345-1348.

Wallace, L. (2016). Using *Google* Web speech as a springboard for identifying personal pronunciation problems. *Proceedings of the 7th Annual Pronunciation in Second Language Learning and Teaching Conference*. Retrieved from https://apling.engl.iastate.edu/alt-content/uploads/2016/08/PSLLT7_July29_2016_B.pdf.

Nguyen, L. T., Newton, J. (2019). Corrective feedback in pronunciation teaching: A Vietnamese perspective. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Ames, IA, September 2018 (pp. 201-212). Ames, IA: Iowa State University.

PRESENTATION/POSTER

CORRECTIVE FEEDBACK IN PRONUNCIATION TEACHING: A VIETNAMESE PERSPECTIVE

Loc Tan Nguyen, University of Economics – Ho Chi Minh City, Vietnam
Jonathan Newton, Victoria University of Wellington, New Zealand

Limited research has investigated teachers' and learners' beliefs about corrective feedback in pronunciation teaching. The current study addresses this gap by examining teachers' and students' perspectives on corrective feedback in pronunciation teaching in an EFL context where it has not hitherto been researched, namely Vietnamese tertiary education. Data included observations and video-recordings of six 90-minute communication classes, and interviews with both teachers and students. Teacher interviews included stimulated recall based on video-recordings of their lessons. Student focus group interviews provided insights into how the students perceived the efficacy of the pronunciation instruction they received. The study highlighted the teachers' stated beliefs about the effectiveness of their practice of primarily delivering corrective feedback through recasts and/or prompts. The students were also able to articulate clear perspectives on corrective feedback in pronunciation teaching, but overall these did not align with those of the teachers. The paper concludes by discussing the implications of this misalignment.

INTRODUCTION

Corrective feedback (CF) refers to teachers responding to learner erroneous utterances (Ellis, 2006). Although CF has been showed to be beneficial for second language learning (Li, 2010; Lyster, Saito, & Sato, 2013), limited research has investigated teachers' and students' beliefs and attitudes towards CF in pronunciation teaching, and no such research of which we are aware has been carried out in the Vietnamese EFL context. Given that millions of teachers and learners are currently teaching and studying English from primary schools to universities throughout Vietnam, it is necessary to look into teachers' and students' perspectives of CF in pronunciation teaching in this EFL context and to examine Vietnamese EFL learners' pronunciation instructional needs.

LITERATURE REVIEW

Research has shown ESL/EFL pronunciation teaching to be typically reactive and unplanned in response to individual student's pronunciation errors, usually in the form of recasts (giving model pronunciations) and/or prompts (encouraging self-correction by giving meta-linguistic clues). For example, Foote, Trofimovich, Collins, and Urzúa (2016) found that ESL teachers in Canada mainly corrected students' pronunciation errors of individual sounds through recasts and/or prompts. Murphy (2011) also found that over 90% of the 36 teacher participants in four different private schools in the Dublin area of Ireland corrected learners' pronunciation errors when they read aloud. A study in Malaysian EFL context also showed that the teachers at a university corrected students' segmental errors through repetition (Wahid & Sulong, 2013). Overall, these research findings revealed that teachers limited their pronunciation instruction to particular types of teaching technique, the most common being CF through repetition.

This approach finds some support in classroom-based studies which have shown CF to be effective in improving learners' pronunciation. For instance, Saito and Lyster (2012) examined the effects of L2 pronunciation instruction with CF on learners' outcomes. Sixty-five adult intermediate Japanese ESL learners in Montreal, Canada were divided into three groups (one control and two experimental) with each group receiving four hours of pronunciation instruction. The two experimental groups worked on the same activities, but one of them additionally received CF in the forms of recasts. Pre-test and post-test results showed that learners who received instruction with CF significantly improved their pronunciation but those without did not. In addition, learners receiving CF also outperformed the control group on similar task items. Positive findings for CF were also reported by Dlaska and Krekeler (2013) in a study involving 169 intermediate adult learners of German. An analysis of post-intervention oral reading by the participants showed that learners who received explicit individual CF from the teacher were found to be easier to comprehend than those who did not. However, neither of these studies investigated teachers' and students' beliefs and attitudes about CF in pronunciation teaching, the topic of the current study.

THE CURRENT STUDY

Research questions

The research addresses the following research questions:

1. How do EFL teachers at a Vietnamese university teach pronunciation?
2. How do the (a) teachers and (b) students perceive the effectiveness of CF in the form of recasts and/or prompts in pronunciation teaching?
3. How do the students expect to be taught pronunciation?

Participants

Six Vietnamese EFL teachers, both male (n = 1) and female (n = 5), participated in the study. The teachers, aged from 29 to 52, were given the pseudonyms 1A, 2B, 3C, 4D, 5E, and 6F for the purpose of this report. All had an MA degree in TESOL (n = 3) or Applied Linguistics (n = 3) and had been teaching at the university from six to 23 years. Twenty-four students (six groups) voluntarily participated in focus group interviews. The students, aged from 19 to 23, included both male (n = 12) and female (n = 12), and had been studying English from seven to 14 years at the time of data collection.

Data collection

Data was collected through classroom observations, stimulated recall (SR) interviews with the teachers, and focus group (FG) interviews with student participants from each of six classes. The classroom observations consisted of non-participant observation of two 45-minute periods for each of the six classes, each taught by a separate teacher (a total of 540 minutes of observation data). The classes were elementary and pre-intermediate level, and covered vocabulary, listening, speaking, and grammar, but not writing. All the observations were audio-video taped, with author 1 also taking unstructured field notes.

One day after each observation, the observed teacher was interviewed for about 30 minutes. In the interviews, each teacher first watched and was asked to comment on two selected excerpts from an audio-video recording of their previous lesson in which pronunciation instruction was present. One except involved a recast and the other a prompt. In the second part of the interview, the teachers were asked general questions about their pedagogic decision-making in relation to pronunciation teaching and how effective they perceived it to be, and why. For the purpose of this paper, we focused on the teachers' answers to the question "Do you think correcting students' errors like this [recasts and/or prompts] is effective and why?"

The student FG interviews each lasted for about 20 minutes each and were audio recorded. The groups first watched the two excerpts of their teacher teaching from the previous class and were given a brief explanation of the nature of the recasting and prompting that were illustrated in these excepts. Then each student was asked in turn to comment on their perception of the value for their pronunciation learning of CF through recasts and prompts. During the interviews, they were also asked to elaborate on how they would like to be taught pronunciation.

Data analysis

A qualitative content analysis approach was adopted for the present study data. This involved an iterative, cyclical and inductive process of identifying and refining themes and categories in the observation and interview data (Duff, 2008). For the purpose of coding the observation data, author 1 adopted Foot et al.'s (2016) four-category scheme to identify and code parts of each lesson where pronunciation instruction was present. These included: (1) *Planning:* pre-planned versus reactive; (2) *Target*: segmental versus supra-segmental; (3) *Specific form* (sound contrast); and (4) *Impact:* involving individual students versus the whole class. Based on classroom-based research by Saito (2011), all instances of the teachers giving CF to students' pronunciation were further coded as recasts or prompts. Another Vietnamese EFL teacher was trained to code a sample of pronunciation teaching episodes from the lessons. A comparison of coding by the two coders showed an agreement percentage of over 98%.

Note that this paper reports on a subset of findings from a larger scale research project on pronunciation teaching at tertiary level in Vietnam. We briefly report on the observational data, but our main focus is on teacher and learner cognition with respect to the specific topic of CF through recasts and/or prompts.

RESULTS

The teachers' pronunciation teaching

The observational data shows that all the teachers only used a reactive focus-on-form approach (Basturkmen, Loewen, & Ellis, 2004) to pronunciation teaching. In other words, the teachers responded to individual students' pronunciation errors through recasts and/or prompts, rather than working from a pre-planned syllabus. The pronunciation teaching episodes identified for each teacher are summarised in Table 1.

Table 1

The teachers' pronunciation teaching episodes

Instructor	**Recasts**	**Prompts**	**Total**
Teacher 1A	Segments: 13 Linking: 2 Word stress: 2	Segments: 6 Linking: 2	Segments: 19 Linking: 4 Word stress: 2
Teacher 2B	Segments: 8 Word stress: 1	Segments: 6; Intonation: 1	Segments: 14 Intonation: 1 Word stress: 1
Teacher 3C	Segments: 4 Linking: 1	Segments: 2	Segments: 6 Linking: 1
Teacher 4D	Segments: 8 Word stress: 2	Segments: 5	Segments: 13 Word stress: 2
Teacher 5E	Segments: 11	Segments: 3 Word stress: 2 Intonation: 1	Segments: 14 Word stress: 2 Intonation: 1
Teacher 6F	Segments: 11 Word stress: 1	Segments: 6 Word stress: 1	Segments: 17 Word stress: 2
All teachers	**64** Segments: 55 Word stress: 6 Linking: 3	**35** Segments: 28 Word stress: 3 Linking: 2 Intonation: 2	**99** Segments: 83 Word stress: 9 Linking: 5 Intonation: 2

As shown in Table 1, a total of 99 pronunciation teaching episodes were identified across the teachers, 83 of which were focused on segmental errors of individual sounds, nine on word stress, five on linking, and two on intonation. Clearly the teachers were most focused on correcting errors in the production of individual sounds at the word level. The following teaching episodes illustrate this.

<u>Episode 1</u>: (*Note*: T = teacher; S = student; Ss = students)

T: Okay. Now, which four adjectives do we use to describe this car?
S1: It's stylist and powerful (*pronounced as* /pɒwefʊl/ *with no lexical stress*).
T: Say 'powerful' (*emphasized lexical stress and vowel production*).
S1: Powerful.
T: That's right. What else? You, please.

<u>Episode 2</u>:

T: When you make your presentation, if you pronounce incorrectly, then you won't get good scores. Beside content, you must pronounce intelligibly for people to understand. Remember? Say these words again for me, please. How do you say this? (*pointing to the first word in the list*)
Ss: Male (*pronounced as* /meʊ/).

T: No. Look at the vowel and the final sound. Say…?
Ss: male.
T: Good. This word? (pointing to the second word in the list)
SS: image (pronounced as /imei/).
T: not 'mei' but…?
Ss: image (*pronounced vowel correctly but dropped the final sound*).
T: Yes. What about the final consonant?
Ss: image.
T: That's right. Now, say 'image'
Ss: image.

This pattern of focusing on segmental errors and of correcting these errors through repetition or awareness raising for self-correction was identical across the observed classes of all six teachers.

We will now turn to examine the teachers' perspectives of CF through recasts and/or prompts in pronunciation teaching.

The teachers' perspectives of CF in pronunciation teaching

In response to the question about the effectiveness of their pronunciation teaching, all the teachers stated that giving CF through recasts and/or prompts was effective. They were then asked to elaborate on why they said so. Four different reasons were given by the teachers as presented in Table 2.

Table 2

The teachers' stated beliefs about the effectiveness of CF in pronunciation teaching

Reasons for giving CF	Frequency	Teacher(s) mentioned
1. Time constraints	n = 6	1A, 2B, 3C, 4D, 5E, 6F
2. Students' errors	n = 6	1A, 2B, 3C, 4D, 5E, 6F
3. Students' awareness	n = 6	1A, 2B, 3C, 4D, 5E, 6F
4. Students' proficiency	n = 2	2B, 5E

Note: n = number of teachers

The first point all the teachers made was that time constraints dictated the approach they took. As Teacher 6F stated,

(1) "*There's been an overload of knowledge in the curriculum [...] Teaching time is too limited but there're so many students in class. There're only four periods each week but what must be taught is too much [...] So I think correcting students' pronunciation errors like this [through recasts and/or prompts] is the most effective way.*" (Teacher 6F)

As shown in this and other comments, all the teachers found that saw that time constraints combined with an overloaded curriculum and large classes all meant that there was no better way to correct students' pronunciation than by providing a model and encouraging self-correction.

Second, all the teachers believed that giving CF in the form of recasts and/or prompts is useful because the errors they correct are common amongst the students. For instance, Teacher 4D explained:

(2) "*As I've just said, there's not enough time to transfer all the content in the curriculum to students, and so correcting students' pronunciation errors like this [through recasts and/or prompts] is the fastest and most effective way. Also, these [pronunciation] errors are common amongst Vietnamese learners, so all the students in class will be aware of the errors and can correct themselves.*" (Teacher 4D)

Overall, the teachers were all willing to teach pronunciation but insisted on staying on schedule in implementing the curriculum. To achieve both these goals, the teachers reported that the focused reactive pronunciation instruction they carried out was effective because it allowed them to balance both these goals.

Third, all the teachers believed that giving CF helped raise learner awareness of pronunciation errors, which was effective in encouraging students to self-correct. As Teacher 3C noted,

(3) "*Correcting a student's pronunciation errors not only works for that student but it also makes all other students aware of such errors and so they can correct themselves. This way could help me save time for other tasks.*" (Teacher 3C)

The belief that raising awareness can help improve learners' pronunciation has been supported by scholars such as Ducate and Lomicka (2009), Kennedy, Blanchet, and Trofimovich (2014), and Ramírez Verdugo (2006). In this EFL context, the teachers believed that pushing students to pay attention to practicing pronunciation leads to improvements. Teacher 1A, for instance, said that if teachers make students aware that pronunciation errors cause misunderstandings and/or communication breakdowns, then their attitudes towards pronunciation will be changed positively and they will pay more attention to practice.

Finally, Teachers 2B and 5E reasoned that since students were at a low level of English proficiency, CF through recasts and/or prompts was effective. The teachers believed that students who are not very good at English benefit from CF such as recasts and/or prompts whereas teaching pronunciation explicitly works more effectively with students of higher proficiency. Teacher 5E, for example, said:

(4) "*[...] Most students of our university are not very good at English. So, I think the best way is to correct their pronunciation errors. This is more or less useful for their pronunciation learning. They'll know where they're mistake and so become more conscious in practicing pronunciation. I think teaching pronunciation explicitly works more effectively with better students [...]*" (Teacher 5E)

In brief, the teachers' stated beliefs show that giving CF through recasts and/or prompts is beneficial to students' pronunciation learning. The following section reports on the students' thoughts and beliefs about CF in the form of recasts and/or prompts in pronunciation teaching.

The students' perspectives of CF in pronunciation teaching and their instructional needs

In response to the question of how effective CF is in pronunciation teaching, all the 24 student participants gave negative responses. Their rating ranged from *not very effective* to *not effective at all* as visualised in Figure 1.

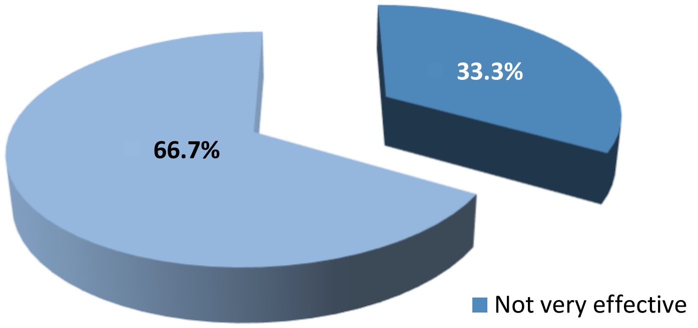


Figure 1. The students' stated beliefs about the effectiveness of CF in pronunciation teaching.

First, Figure 1 shows that one third of the students (n = 8) reported that giving CF was not a very effective approach to pronunciation teaching. Student 3 from FG5, for instance, explained:

(5) "*[...] What we only do is to listen and repeat after the teacher like a machine. Later on, we'll forget all about it because we aren't taught pronunciation theory and don't have opportunities for communication practice either.*" (Student 3, FG5)

This and other extracts show that the students saw teachers correcting pronunciation errors to be a temporary solution and so believed they did not benefit much from it. From the students' responses it appears that they sought instruction that could bring about more long-term effects to their pronunciation skill. Also implied in the students' comments were their expectations of more explicit teaching of pronunciation along with opportunities for communication practice.

Second, two thirds of the students (n = 16) stated that CF through recasts and/or prompts as their teachers did was completely ineffective. According to the students, repeating model pronunciation was too mechanical and difficult for them to remember, and so not at all beneficial to their learning. They believed that their pronunciation could not be improved through this teaching strategy. The following comments illustrate this collective view:

(6) "*I think it's completely ineffective. Teachers' correction of our errors through repetition is too mechanical. I think pronunciation requires time but teachers don't seem to care about it and so they don't spend time teaching pronunciation explicitly in class.*" (Student 3, FG3)

(7) "*It's not effective at all. Repetition drills are just like the way of teaching a parrot how to speak. I can only pronounce the words that teachers have taught. When seeing a new word, I don't know how to read it correctly. So, I don't think my pronunciation improves through this approach.*" (Student 4, FG6)

Overall, the students showed negative attitudes towards CF through repetition in pronunciation teaching. They reported finding it mechanical and thus not beneficial to their pronunciation skill. What the students said they expected was that teachers take better care of their pronunciation skill by spending more time teaching pronunciation explicitly in class.

The question, then, is how the students expected to be taught pronunciation and why. In response, all the students stated that they wanted pronunciation to be taught in such a way that can improve not only their pronunciation but also listening and speaking skills. The following comment is representative:

(8) "*I don't know how to say but I expect teachers to teach pronunciation in such a way that provides us with more communication practice. The way that helps me improve pronunciation, listening, and speaking skills at the same time [...]*" (Student 1, FG1)

As mentioned above, although the students did not use the term "the communicative teaching of pronunciation", it was implied in their responses that they believed that teaching pronunciation communicatively not only helps improve their pronunciation but also communication skills.

Elaborating on why they wanted to be taught pronunciation communicatively, the students said it is better than the approach that bases itself mostly on repetition. According to the students, if pronunciation is taught communicatively, learners' listening and speaking skills improve (n = 18), classroom tension reduces and learner motivation increases (n = 12), learner comprehension of teacher instruction speeds up (n = 7), and interactions are promoted (n = 4). Figure 2 provides a visual representation of the students' stated beliefs about the communicative teaching of pronunciation.

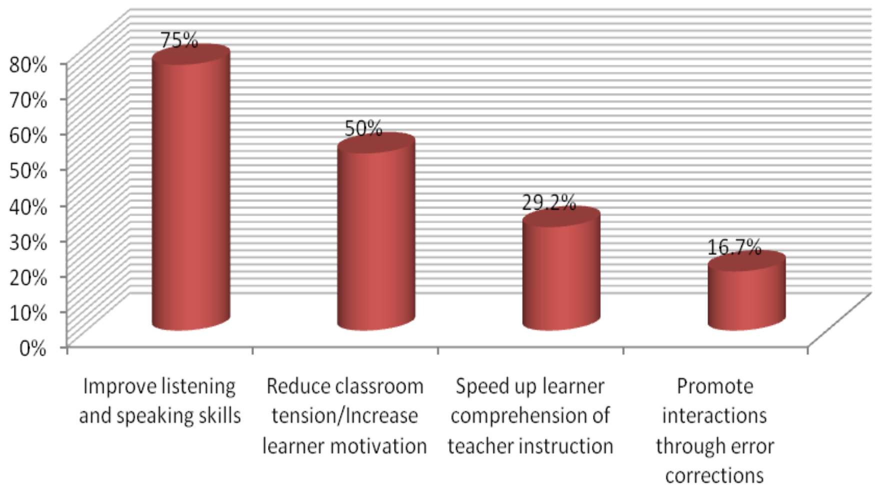


Figure 2. The students' stated beliefs about communicative pronunciation teaching.

First, 18/24 of the students believed that it is more practical for pronunciation to be taught communicatively so their listening and speaking skills can simultaneously improve. As the students saw it, the communicative teaching of pronunciation provides more opportunities and allows them to practice pronunciation through communication situations. The following extracts are representative:

(9) "*Pronunciation should be taught communicatively. This way is more practical because we can apply what we've learned in communication. And when we have more opportunities for practice, our listening and speaking skills will be improved.*" (Student 1, FG6)

(10) "*I'd prefer teachers to teach pronunciation communicatively because it's practical and more interesting. Learning through repetition is very boring, and we'll forget everything quickly. When teachers give us more communication practice, our listening and speaking skills can improve.*" (Student 2, FG4)

Second, 12/24 of the students stated that the communicative teaching of pronunciation helps reduce classroom tension and increase learner motivation. According to the students, when pronunciation is taught communicatively, the classroom atmosphere is more interesting and welcoming. Thus, learners will be more motivated in coming to class. They said:

(11) "*I expect teachers to teach pronunciation communicatively because students will have more opportunities for practice in communication situations. Also, the classroom atmosphere will be more interesting, making students more motivated coming to class.*" (Student 2, FG2)

The students' primary concern in characterising their preferred approach to pronunciation teaching was the classroom atmosphere. They believed that if classes are interesting, then they will be motivated to attend. Moreover, they will be more active in class and thus teacher's instruction is more beneficial to their learning as they become more productive learners.

Third, seven students also believed that the communicative teaching of pronunciation speeds up learner comprehension of teacher instruction. They said teacher instruction is more comprehensible this way and thus they can absorb the knowledge faster and the outcomes will be better. For example:

(12) "*I'd prefer pronunciation to be taught communicatively because it makes me more interested in learning. The lessons will be more comprehensible and I can absorb the knowledge transferred by my teacher more quickly. So, the results will be much better.*" (Student 3, FG3)

Finally, teaching pronunciation communicatively also promotes interactions in the classroom through peer and teacher corrections as articulated by four of the students. They commented:

(13) "*I think pronunciation should be taught more effectively. Repetition drills are boring, mechanical, and not practical. But when teaching [pronunciation] communicatively [...] there'll be more interactions between teachers and students. We can correct each other and teachers correct our errors too [...]*" (Student 3, FG5)

This and similar comments show the students' belief about how the communicative teaching of pronunciation encourages interactions in class. They reported that communication practice helps them identify their own pronunciation problems which need correction from peers and the instructor. In this sense, interactions amongst pair/group members and between the teacher and students will be facilitated.

In sum, the students did not value CF through recasts and/or prompts. Instead, they expressed a strong need for more explicit communicative teaching of pronunciation.

DISCUSSION AND CONCLUSION

The study found that the teachers' pronunciation teaching was restricted to error correction through repetition. Their stated beliefs show that they believe this is an effective way to address pronunciation in their English classes. The value the teachers put on error correction finds support from a general claim that CF is beneficial for improving pronunciation (Lyster et al., 2013; Saito & Lyster, 2012). However, Foote et al. (2016) have argued that instructors can be over-reliant on CF and this fails to address pronunciation proactively. They claim that without explicit instruction that first helps students understand a target feature, the feedback teachers give is less likely to beneficial to student learning. If this is the case, then our finding that the teachers relied in CF is not encouraging.

The study has also found that the students were not in favour of CF as a pronunciation teaching approach and considered it as unhelpful to their learning. Although L2 pronunciation acquisition can be facilitated by repetition drills (Trofimovich & Gatbonton, 2006), the students found these drills too mechanical, boring and ineffective and so expected pronunciation to be taught communicatively so that they can improve not only pronunciation but also listening and speaking skills. This approach to pronunciation teaching has been supported by scholars such as Isaacs (2009), Celce-Murcia, Brinton, and Goodwin (2010), and Avery and Ehrlich (2013). Spada and Lightbown (2008) have argued that communicative activities may be the best choice for learners to develop fluency and automaticity necessary for oral communication outside the classroom. In this EFL context, although the students have little need to use English for oral interactions outside the classroom, their favourable attitudes towards the communicative teaching of pronunciation are encouraging. The value that the students put on this teaching approach is consistent with learners in an American ESL setting who were reported to want more real-life communication situations to practice the target pronunciation features (Vitanova & Miller, 2002).

Perhaps the most important aspect of our findings is the dissonance between the views of the teachers and students on the efficacy of current pronunciation teaching practices. Regardless of which views find the most support in the research literature or which approach is the most effective and realistic, this result suggests that there is room for dialogue between teachers and students so that each can gain greater understanding of the views of the other and modify practices and expectations accordingly.

ABOUT THE AUTHORS

Loc Tan Nguyen has been an EFL instructor at the University of Economics, HCMC, Vietnam for 15 years. He is currently a PhD candidate in Applied Linguistics at Victoria University of Wellington, New Zealand. His research interests arelanguage teaching methodology, classroom-based pronunciation teaching, and teacher professional learning.
Correspondence: tanloc_nt@ueh.edu.vn

Jonathan Newton is an Associate Professor in Applied Linguistics at Victoria University of Wellington, New Zealand. His research interest is language teaching methodology and task-based language teaching (TBLT), teacher cognition and professional learning for language

teachers, the interface of culture and language in language teaching and learning/communication training and materials design for the multicultural workplace.

Correspondence: jonathan.newton@vuw.ac.nz

REFERENCES

Avery, P., & Ehrlich, S. (2013). *Teaching American English pronunciation*. Oxford: Oxford University Press.

Basturkmen, H., Loewen, S., & Ellis, R. (2004). Teachers' stated beliefs about incidental focus on form and their classroom practices. *Applied Linguistics*, *25*(2), 243-272.

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

Dlaska, A., & Krekeler, C. (2013). The short-term effects of individual corrective feedback on L2 pronunciation. *System*, *41*(1), 25-37.

Ducate, L., & Lomicka, L. (2009). Podcasting: An effective tool for honing language students' pronunciation? *Language Learning & Technology*, *13*(3), 66.

Duff, P. (2008). *Case study research in applied linguistics*. New York: Routledge.

Ellis, R. (2006). Researching the effects of form-focussed instruction on L2 acquisition. *AILA review*, *19*(1), 18-41.

Foote, J. A., Trofimovich, P., Collins, L., & Urzúa, F. S. (2016). Pronunciation teaching practices in communicative second language classes. *The Language Learning Journal*, *44*(2), 181-196.

Isaacs, T. (2009). Integrating form and meaning in L2 pronunciation instruction. *TESL Canada Journal*, *27*(1), 1-12.

Kennedy, S., Blanchet, J., & Trofimovich, P. (2014). Learner pronunciation, awareness, and instruction in French as a second language. *Foreign Language Annals*, *47*(1), 79-96.

Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning*, *60*(2), 309-365.

Lyster, R., Saito, K., & Sato, M. (2013). Oral corrective feedback in second language classrooms. *Language Teaching*, *46*(1), 1-40.

Murphy, D. (2011). An investigation of English pronunciation teaching in Ireland. *English Today*, *27*(4), 10-18.

Ramírez Verdugo, D. (2006). A study of intonation awareness and learning in non-native speakers of English. *Language Awareness*, *15*(3), 141-159.

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.

Saito, K., & Lyster, R. (2012). Effects of form-focused instruction and corrective feedback on L2 pronunciation development of/ɹ/by Japanese learners of English. *Language learning, 62*(2), 595-633.

Spada, N., & Lightbown, P. M. (2008). Form-focused instruction: Isolated or integrated? *TESOL Quarterly, 42*(2), 181-207.

Trofimovich, P., & Gatbonton, E. (2006). Repetition and focus on form in processing L2 Spanish words: Implications for pronunciation instruction. *The Modern Language Journal, 90*(4), 519-535.

Vitanova, G., & Miller, A. (2002). Reflective practice in pronunciation learning. *The Internet TESL Journal, 8*(1).

Wahid, R., & Sulong, S. (2013). The Gap between research and practice in the teaching of English pronunciation: Insights from teachers' beliefs and practices. *World Applied Sciences Journal, 21*, 133-142.

Scott, J. H. G. (2019). Who follows the rules? Differential robustness of phonological principles. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Ames, IA, September 2018 (pp. 213-225). Ames, IA: Iowa State University.

PRESENTATION/POSTER

WHO FOLLOWS THE RULES? DIFFERENTIAL ROBUSTNESS OF PHONOLOGICAL PRINCIPLES

John H. G. Scott, Indiana University

This study investigates sensitivity to violations of two phonological rules by 14 native speakers of German and 23 L2 learners (L1 American English). Participants completed a phoneme detection task, listening for [t] in pseudowords, including sequences that conformed to the German rule of Dorsal Fricative Assimilation (e.g., [b**ax**t]/[b**ɛç**t]) or violated it (e.g., *[b**aç**t]/*[b**ɛx**t]). Additional stimuli included [h] in initial position (e.g., [**h**amt]), where it is legal in English and German, and in syllable codas (e.g., *[ba**h**t]/*[bɛ**h**t]), where it is banned in both. Systematic reaction time effects in response to phonotactic violations are analyzed as evidence that the principle is psychologically real in the phonological grammar. Learners exhibited reaction time effects for both types of violations. In contrast to previous findings, Germans showed no effect for violation of Dorsal Fricative Assimilation; they also showed a slowdown trend for [h] in coda position. These findings suggest important differences between phonological knowledge types and between native and L2 learner exposure to phonological principles.

INTRODUCTION

Second language (L2) phonology (i.e., systems of sounds and their abstract relationships) is necessarily intermingled with phonetics (physical properties of speech sounds). Phones are the lowest common denominator between first language (L1) and L2 in perception and production, and thus a necessary element in L2 phonology studies. Therefore, L2 phonology research must include both phonetic sensitivity and phonological knowledge at different stages of development. Focusing on German native speakers (NS) and early L2 learners (L1 American English), this study investigates the psychological reality of two phonological rules—namely, the German alternation of the palatal fricative [ç] with the velar fricative [x] (Dorsal Fricative Assimilation: DFA) and the restriction in both English and German that bans [h] from occurring at the end of syllables.

LITERATURE REVIEW

Dorsal fricative assimilation

Palatal [ç] and velar [x] fricatives are novel for L1 American English speakers. In German, they are called *ich-Laut* ('I-sound' after the pronoun *ich* 'I') and *ach-Laut* ('oh-sound' after the interjection *ach* 'oh'). In the phonological literature, they famously alternate in word-internal post-vocalic position, depending on whether they follow a front vowel or a back vowel (e.g., *Buch*, [buːx] 'book, SG.' vs. *Bücher*, [ˈbyː.çɐ], 'book, PL.') or certain consonants (e.g., *Mönch*, [mœnç], 'monk'). In general terms, *ach*-Laut surfaces as velar [x] after non-low back tense vowels, as uvular [χ] after low vowels, and as either [χ] or [x] after non-high back lax vowels [ʊ/ɔ] (Wiese, 1996, pp. 209–210). Some exceptions exist, especially with the invariant diminutive suffix (e.g., *Kuh*, 'cow,' [kuː] vs. *Kuhchen*, 'cow, DIM.,' [ˈkuː-çən]). I treat DFA as an active allophonic alternation in German, whereby, at least in monosyllables, [x]/[χ] surfaces after back vowels and [ç] surfaces elsewhere. For convenience, sequences that conform to DFA are termed "Good;" violations of DFA are "Bad."

Distribution of /h/

The feature [+spread glottis] aligns [h] with voiceless aspirated stops (e.g., [p^h]; Halle & Stevens, 1971). This natural class makes the prediction that, in some language(s), some phonological rule or constraint applies to or is triggered by both types of segments. American English bears this out: aspirated stops and /h/ pattern together. For a thorough treatment of the distribution of /h/ and aspirated stops with respect to the prosodic foot in English, see Davis and Cho (2003). Jessen (1998, pp. 152–153) notes the same alignment of /h/ and aspirated stops for German, using [tense] as the operative feature. Here, the fact that [h] may not appear in the right syllable margin is called the *Coda-[h] ban. To unify terminology between conditions, items in [h]-conditions that have [h] in the left syllable margin are called "Good;" items with [h] in the right margin, violating the *Coda-[h] ban, are "Bad."

Phoneme detection

Figure 1 presents a summary of phoneme detection studies that have investigated phonological knowledge of place assimilations. Otake, Yoneyama, Cutler, and Van der Lugt (1996) investigated Japanese regressive (right-to-left) nasal assimilation to the place of a following consonant, where place assimilation is obligatory, and Weber (2001a, 2001b) replicated this. Weber (2001a, 2001b, 2002) also investigated sensitivity to violations of progressive (left-to-right) DFA in German. Otake et al. (1996) and Weber (2001a, 2001b) found consistent and strong reaction time (RT) slowdown (inhibition) when nasals and following stops clashed in place of articulation ("Bad"). Weber instructed participants to listen for dorsal fricative allophones—velar [x] or palatal [ç]--and press a button when they were heard. Native German speakers responded slightly faster (facilitation) for front+[x] "Bad" sequences, but not for back+[ç] (also "Bad"). Weber argues that this is because this type of sequence occurs in German (e.g., *Kuh-chen*). Thus, the alternation of [ç] and [x], governed by the place of the preceding segment, seems to have psychological reality for German NSs. Lindsey's (2013) replication included L2 German learners (L1 American English), finding strong and consistent RT inhibition in both German NSs and advanced L2 learners for violations of DFA (all "Bad" sequences). These results suggest that learners can acquire sensitivity to novel L2 alternations. It is unclear whether and under what conditions "Bad" sequences inhibit or facilitate phonological processing. In contrast to the studies cited here, this study replicates and expands on Weber's and Lindsey's studies to include the *Coda-[h] ban, but avoids cross-language labels by using the familiar listening target [t].

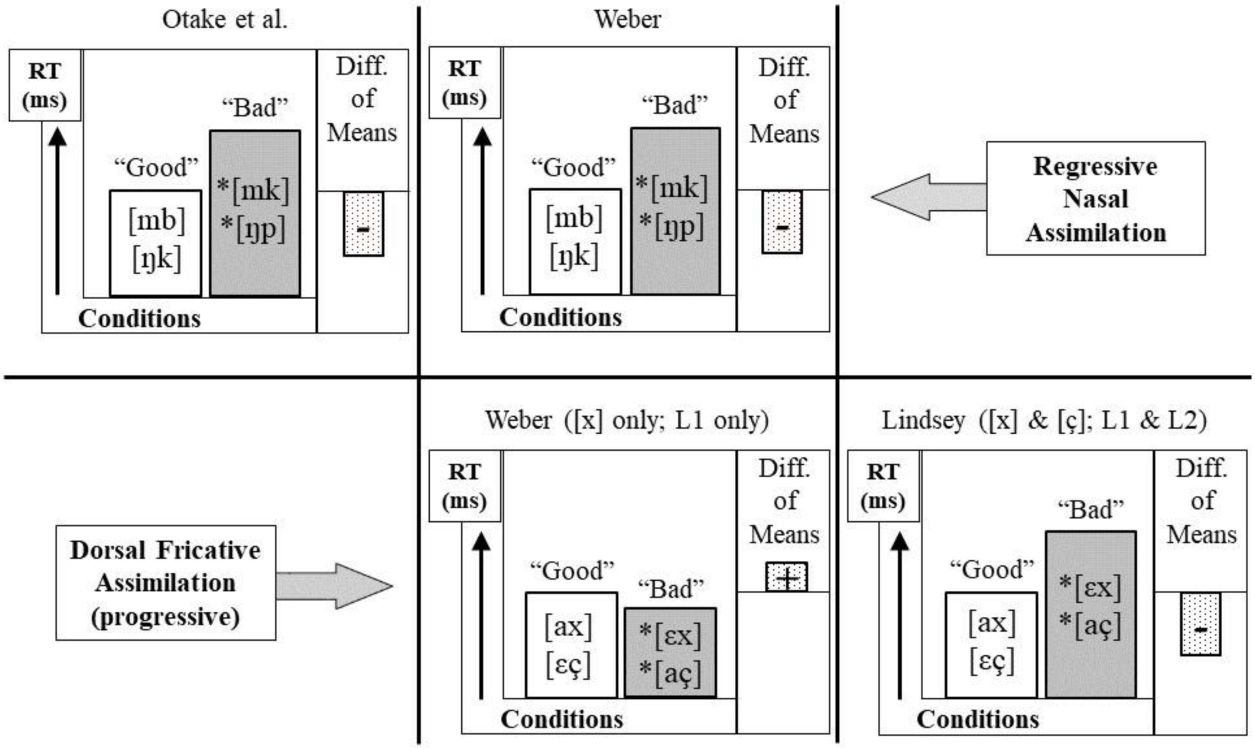


Figure 1. Summary of phoneme detection studies investigating place assimilations. Diagrams show examples for reported patterns in these conditions, not actual data. Difference of means represents mean RT of "Bad" conditions subtracted from "Good."

Research questions

The current investigation pursues the following research questions:

1. What RT effects arise from violation of progressive (left-to-right) assimilation expectations?
2. What RT effects arise from violation of syllable constraints on phoneme distribution?
3. How robust (consistent and strong) are these phonological principles compared to other (obligatory) assimilation rules?

If DFA is acquired as phonological knowledge, then "Bad" sequences *[aç/ɛx] that violate DFA should trigger a RT effect compared to expected sequences [ax/ɛç] (RQ 1). The restricted distribution of [h] in English and German represents a different type of phonological knowledge from rules of assimilation. When [h] appears illegally in the right syllable margin, NSs and L2 learners should exhibit consistent and strong inhibition (RQ 2). The *Coda-[h] ban in the right syllable margin (e.g., *[ba**h**t]) is without exception in English and German. Similarly, if allophonic alternation from DFA is acquired as an obligatory phonological rule, then violation of DFA should yield a similarly strong and consistent effect (RQ 3).

METHOD

Participants

Participants included 14 German NSs (2 male, 12 female; 20–29 years old, mean age = 22.9) in Stuttgart, Germany [i] and 23 NSs of American English (12 male, 11 female; 18–23 years old; mean age = 19.6) enrolled in second-semester German at a large Midwestern university. At the time of data collection, German L2 learners reported 1–11 previous semesters of secondary or post-secondary instruction in German; 12 reported just one previous semester and seven reported six or more previous semesters. Participants 2002 and 2039, respectively, reported initial German exposure at age 3 or "very young," but no use of German until the late teens. (Results of both contribute to the group majority trends reported here.) Data collection with L2 learners was conducted at two time points during the same spring semester. Six participants completed the task at both time points; only their data from the later time point were analyzed.

Materials

The experiment included 384 nonword trials with [t] as the listening target. Critical trials (*n* = 48) were balanced for "Good" and "Bad" in three pairs of conditions (*n* = 8 trials / condition), shown in Table 1. "Good" conditions included back Match (e.g., [gl**ax**t]), front Match (e.g., [gl**ɛç**t]), and Onset-[h] (e.g., [**h**amt], [**h**ɛlkt]); their "Bad" counterparts were back-front Mismatch (e.g., *[gl**aç**t]), front-back Mismatch (e.g., *[gl**ɛx**t]), and Coda-[h] (e.g., *[gla**h**t], *[klɛ**h**t]). There were 144 trials containing [t] in other (nonfinal) positions, and all 192 [t]-trials were balanced by 192 fillers without [t].

Table 1

Onset and duration of pre-targets and listening targets (means and standard deviations)

		Fricative				Listening Target [t]			
		Onset (ms)		Duration (ms)		Onset (ms)		Duration (ms)	
Condition	*n*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Licit position									
Match [ax]	8	258	61	162	19	420	53	184	14
Match [ɛç]	8	275	67	160	20	435	76	180	18
Onset [h]	8	5	1	114	20	393	55	176	28
Illicit position									
Mismatch *[aç]	8	281	69	199	20	480	53	149	13
Mismatch *[ɛx]	8	300	60	153	17	454	51	180	20
Coda Cluster *[h]	8	236	65	316	39	552	73	200	24
Filler with [t]	144	-	-	-	-	102	109	81	44

Note. For [x]- and [ç]-conditions, fricatives immediately precede the listening target [t]. For [h]-conditions, "Bad" [h] immediately preceding the listening target in coda clusters is longer than the other fricatives, including "Good" [h] in onsets.

For each item, at least three tokens were recorded in a sound-attenuated booth at a sampling rate of 44100 Hz by a phonetically trained female NS of German from Saxony. The researcher selected tokens on the basis of recording quality. In addition to tokens recorded for experimental trials, six nonword training trials were used: *Tiesel, gamisch, frettig, Skirm, Prasen,* and *Schloft*. In the training phase, these were presented as one block in the order shown. Practice trials alternated with training scripts highlighting that [t] occured in various positions.

Procedure

Data collection sessions (90–120 minutes) included a language background questionnaire in participants' native language and two other experiments not reported here. Participants received instructions in their native language verbally and on screen. Stimuli were presented via headphones with volume control. Participants were instructed to listen for the target sound [t] and respond as fast as possible with the space bar when they heard it. As /t/ is phonemic in both English and German, it was readily available to both groups for labeling the listening target. All "Good" conditions ([εç]-Match, [ax]-Match, Onset-[h]), "Bad" conditions (*[aç]-Mismatch, *[εx]-Mismatch, *Coda-[h]), and fillers were randomized in a single block with self-paced breaks after every 32 trials.

Native speakers. German NSs completed the experiment in Stuttgart, Germany. All units ran Windows 7 Professional (Service Pack 1, 64-bit). The experiment was run in OpenSesame (Version 2.9). Participants received €15 cash.

L2 learners. Data were collected from the L2 group at the middle and end of the same semester of their second-semester university German course. Testing was administered by desktop computer running Windows 7 Service Pack 1 (64-bit). Additional specifications varied by computer in the laboratory. The experiment was run in OpenSesame (Version 2.8). Participants received US$10 cash for the first session and 1% bonus German course credit for the second. Returning participants at the second session were entered in a drawing for a $50 cash prize (one per 10 returning participants).

RESULTS

Data preparation

Dependent variables. Trials had a "go" versus "no-go" response format; therefore, only hits (i.e., accurate identification of the listening target present in the stimulus) and false alarms (i.e., inaccurate indication of the listening target's presence in a stimulus that did not contain it) resulted in recorded responses. Only hits were analyzed. Slower RT, a measure of variation in speech processing, is attributed to greater processing load (Weber, 2001b, p. 12). The dependent variable for analysis was derived from the recorded RT.

Derivation of augmented RT. Reaction time measurements logged by OpenSesame were augmented by duration (ms) from the onset of the [t] listening target to the end of the audio file, measured with Praat (Version 6.0.19). Augmented RT was used for analysis. The onsets and durations for critical condition stimuli shown in Table 1 (see Materials) only reflect occurrences of [t] in final position, whereas measurements of [t] presented for fillers collapse instances of [t] in all nonfinal positions. Descriptive statistics in Table 1 are for overview only; the precise onset time for each stimulus was used to calculate the adjusted RT for analysis.

Exclusion criteria. To mitigate the impact of nonparticipation on the analysis, participants with fewer than five hit responses in any of the four conditions were excluded. This criterion excluded nine of the L2 group and five of the NS group from subsequent analyses, retaining data from 14 learners and nine NSs. Table 2 summarizes the data retained for analysis.

Table 2

Data set totals after participant exclusion

	Dorsal Fricatives		Glottal Fricative		
	"Good"	"Bad"	"Good"	"Bad"	
Group	[axt]$_\sigma$ / [ɛçt]$_\sigma$	*[açt]$_\sigma$ / *[ɛxt]$_\sigma$	[hVCt]$_\sigma$	*[…Vht]$_\sigma$	Totals
L2 (*n* = 14)	194	195	100	97	586
NS (*n* = 9)	135	123	60	56	374
Totals	329	318	160	153	*N* = 960

Note. Fillers are excluded from the table and analysis. All critical trials were monosyllables with the listening target [t] in syllable-final position, indicated by]$_\sigma$. For "Good" glottal fricative trials, the penultimate consonant was always licit in that position in both English and German.

Group means

Each participant's mean RT was computed across items for each fricative condition (Dorsal, Glottal) in combination with the factor Context ("Good," "Bad"). Table 3 displays the mean RT for each group by condition. For dorsal fricatives [ç/x], RT in the "Good" condition is equivalent, establishing baseline performance for both groups. The groups differ in their performance with "Bad" dorsal fricatives: NSs' RT is equivalent to the "Good" condition, whereas L2 learners respond more quickly to "Bad" than to "Good." For the glottal fricative [h], NSs are slower than L2 learners in both "Good" and "Bad" conditions. Both groups are slower with "Bad" *Coda-[h] than "Good" Onset-[h], but this is more pronounced for learners.

Table 3

Mean RT (ms) and standard deviation (SD) for each group and each condition

	Native Speakers (N = 9)		L2 Learners (N = 14)	
Condition	RT	*SD*	RT	*SD*
Dorsal Fricatives [ç x]				
"Good" [axt]/[εçt]$_σ$	554	83.6	548	75.2
"Bad" *[açt]/[εxt]$_σ$	551	80.0	521	62.0
Glottal Fricative [h]				
"Good" $_σ$[hVCt]	622	77.5	545	63.9
"Bad" *[…Vht]$_σ$	669	102.9	632	57.2

Analysis of variance

For comparability with earlier phoneme detection studies, a one-way repeated measures analysis of variance (ANOVA) was conducted in SPSS 25 on the RT means for each group and each fricative type, declaring the factor Context (two levels).

Native speakers. There was no main effect of Context on RT in dorsal fricatives [ç/x], $F(1, 8) = 0.038$, $p = .851$, $\eta_p^2 = .005$. This result is unsurprising given the difference of means in this condition was only 3 ms. There was also no main effect of Context found on RT in glottal [h], but there was a trend for slower RT with "Bad" *Coda-[h] (669 ms) than with "Good" Onset-[h] (622 ms) in this group, $F(1, 8) = 2.866$, $p = .129$, $\eta_p^2 = .264$; the partial eta-squared measure of effect size indicates a "would-be" large effect, so this result may have resulted from limited statistical power due to sample size.

L2 learners. The ANOVA showed a main effect of Context on RT in dorsal fricatives [ç/x], $F(1, 13) = 6.874$, $p = .021$, $\eta_p^2 = .346$. This means that "Bad" dorsal fricatives *[aç/εx] have a reliably faster RT (facilitation) than "Good" dorsal fricatives [ax/εç] for this group, and that is a large effect despite the apparently small difference of means (26 ms). This particular effect may not be as robust as the ANOVA suggests. Scott (2019) treats the same data with a mixed effects model, finding only a marginal trend. A main effect of Context on RT was also found in glottal [h], $F(1, 13) = 27.858$, $p < .001$, $\eta_p^2 = .682$, which means that "Bad" *Coda-[h] reliably had considerably slower RT (inhibition) than "Good" Onset-[h], with a large effect size.

Summary of ANOVA. Surprisingly, German NSs do not manifest expected RT shifts for violations of DFA, and they exhibit only a trend for slower RT (inhibition) with "Bad" *Coda-[h] compared to "Good" Onset-[h]. In contrast, L2 learners exhibit RT facilitation with "Bad" dorsal fricatives and marked RT inhibition with “Bad” *Coda-[h], as opposed to “Good” Onset-[h] (RQ 1). Learners react strongly to violation of the *Coda-[h] ban; the NS group shows a similar nonsignificant trend (RQ 2). Thus, L2 learners exhibit more reliable RT effects for both types of violations than NSs do.

Effect consistency and strength

This section investigates observed RT effects within groups by individual differences of means between “Good” and “Bad” conditions for each fricative type.

Subtracting the “Bad” mean RT from “Good” for each NS participant yields categorically negative differences of means for the [h]-conditions (Figure 2). Every L2 learner exhibits a slower mean RT for the “Bad” *Coda-[h] condition. This indicates a consistent and strong inhibition by L2 learners in response to violation of the ban. This is not surprising, as this ban holds for English—it is not new phonological knowledge.

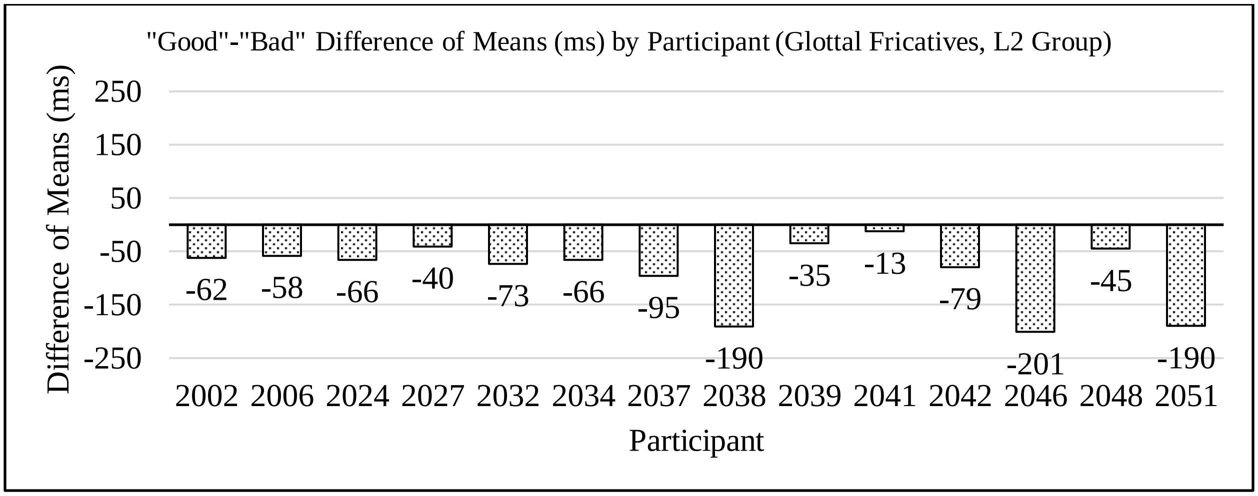


Figure 2. Individual difference of mean RT by participant for [h]-conditions.

Differences of means for the dorsal fricative conditions (Figure 3) are mainly positive, indicating that mean RT for “Bad” dorsal fricatives *[aç/εx] tended to be faster than for the “Good” dorsal fricatives [ax/εç]. However, several L2 learners exhibit no RT effect, and variation is high. Although some L2 learners exhibit strong facilitation in response to violation of DFA, this is not a consistent group effect.

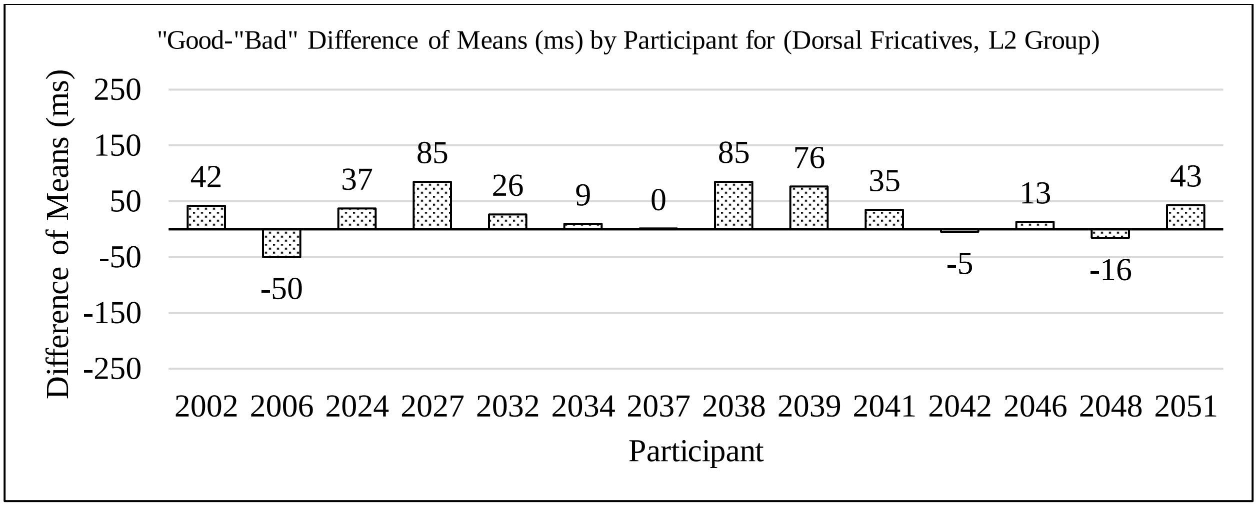


Figure 3. Individual differences of mean RT by participant for dorsal fricative conditions.

Figure 4 displays differences of means between [h]-conditions for NSs. Several NSs show strong inhibition with "Bad" *Coda-[h], several show only minor differences, and variation is high. Two-thirds of these NSs exhibit strong RT shifts, but no single effect is consistent for the group. This suggests that individual Germans differ as to whether the *Coda-[h] ban is psychologically real in their phonological grammars and how it is represented; however, this may be the result of interaction with the preceding vowel (see Discussion).

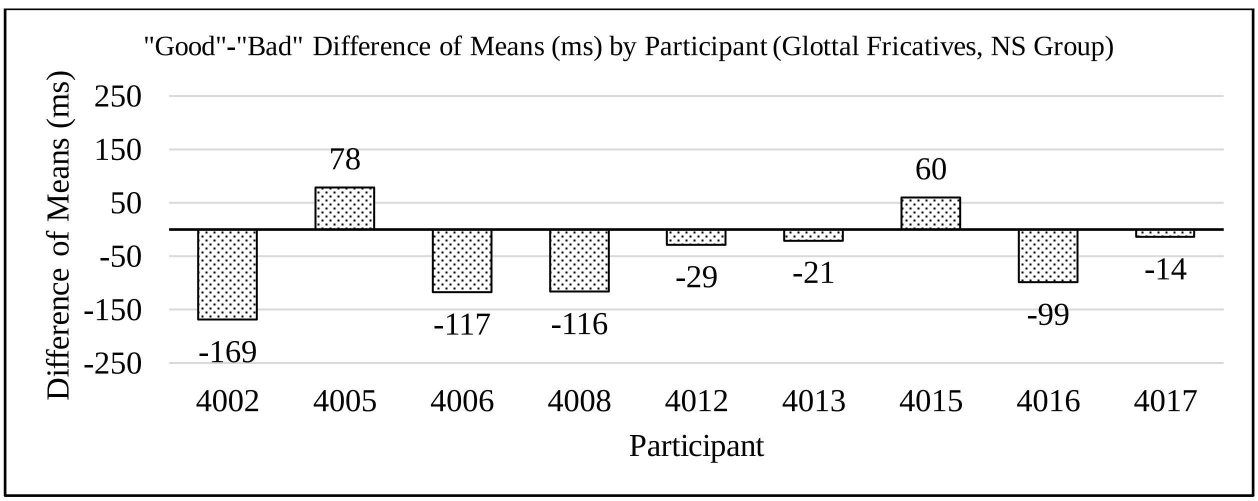


Figure 4. Individual differences of mean RT by participant for [h]-conditions.

Finally, no effect of "Good" versus "Bad" sequence on RT was found in differences of means for NSs in the dorsal fricative conditions (Figure 5). Each participant exhibits weak RT shifts, if any, and differences are not consistently positive or negative. This suggests that individual Germans differ in terms of how (and whether) DFA is represented in their phonological grammars (see Discussion).

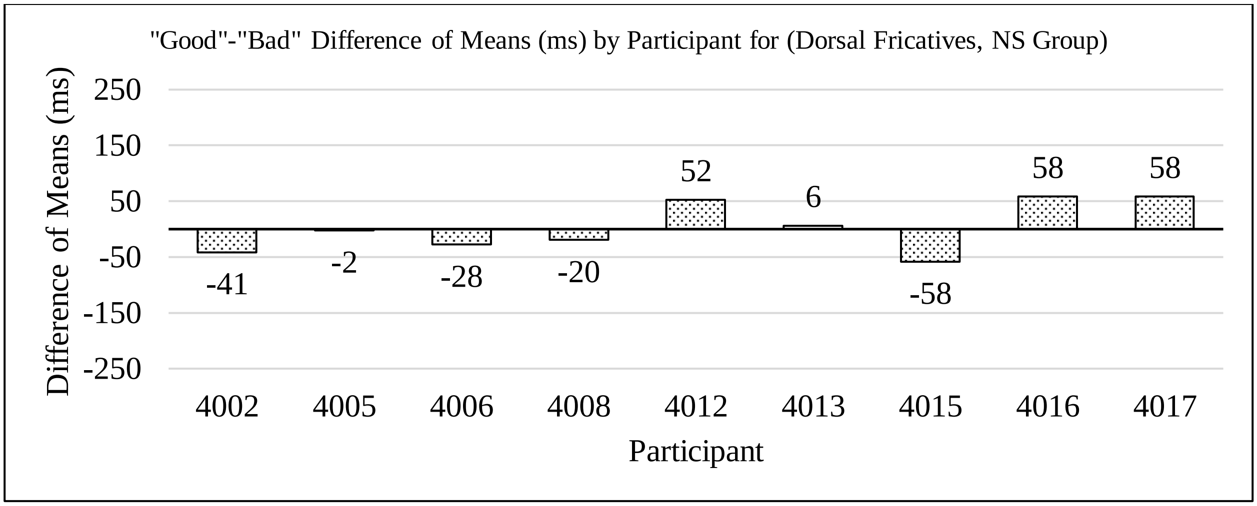


Figure 5. Individual differences of mean RT by participant for dorsal fricative conditions.

General summary

These results support the following conclusions with regard to the Research Questions:

1. Early L2 learners respond faster to violations of German DFA than to place-assimilated sequences. Surprisingly, German NSs manifest no corresponding RT effect.
2. Early L2 learners react more slowly to violations of *Coda-[h] than to trials with licit Onset-[h]. Although predicted to react similarly due to the same ban in German, NSs show only a nonsignificant trend.
3. Learners' slower response to "Bad" *Coda-[h] is robust (strong and consistent). German NSs exhibit high variation in RT in response to these violations—that is, the group is not consistent. Some L2 learners respond faster to "Bad" *[aç/ɛx], but the group is not consistent. Surprisingly, German NSs exhibit no clear RT shifts in response to DFA violations.

DISCUSSION

This study investigated the psychological reality of DFA, which governs the alternation of allophones [ç] and [x] in German, and the ban of [h] in syllable codas in English and German. It adopts the assumption that violation of a phonological rule or constraint increases processing load, which manifests as measurable RT shift. The results suggest that early L2 German learners attend to violation of both phonological rules more than NSs, despite the understanding that both are active in German.

Models of L2 phonological acquisition generally agree that L2 learners are more attentive to subphonemic (phonetic) detail than NSs are. Thus, German NSs may attend less to the wrong allophone appearing in a given DFA context if the allophones themselves are not the listening target. An alternative explanation for Germans' less consistent reactions to DFA violations may lie in the difference between the highly variable linguistic/dialectal exposure of NSs and the low-variability exposure afforded L2 learners in FLA instructed settings (Barriuso & Hayes-Harb, 2018). Phonetic realization of the German dorsal fricative varies widely across German dialects (Hall, 2014); thus, German NSs encounter high variability in the allophones of the dorsal fricative and numerous talkers. In contrast, L2 learners of German as a foreign language rely primarily on instructors for phonological input. As (some) students learn to attend to DFA in perception and production, less variable input may lead to RT effects with “Bad” *[aç/ɛx].

The difference between learner and NS reactions with the *Coda-[h] ban requires a different explanation. German NSs showed a RT slowdown trend in response to “Bad” *Coda-[h]. Exploratory *t*-tests suggest that this trend interacts with the preceding vowel (Scott, 2019). The “Bad” *[ɛh] subcondition shows a significant slowdown compared to its “Good” [ɛç] counterpart, but the “Bad” [ah] does not result in significantly slower mean RT than the “Good” [ax]. This may find explanation in the acoustic similarity of glottal [h] to velar [x], which is legal in that position. Segui, Frauenfelder, and Hallé (2001) describe three types of *phonotactic assimilation*, by which listeners may perceptually ignore a phonotactically illegal phone, or substitute it perceptually with a phone that is legal in that position. German NSs have [x] in their inventory, so may reinterpret acoustic [h] after back vowels (e.g., [a]) as [x] early in processing, then parse it as phonotactically legal [x], triggering no RT effect. However, if L2 learners have not fully acquired (automatized) a fricative that is legal in this position, this mechanism would not be available (cf. Selective Perception Routines; Strange 2011). This experiment included eight “Bad” *Coda-[h] trials: four included [a] and [ɛ], respectively. Of those, some data were missing due to lack of response, leaving less than four data points for each vowel subcondition. The present data set is too small to test this asymmetry without relying on multiple *t*-tests, increasing likelihood of Type I error.

The present study is limited by power and high variation. Although variation among students of foreign language is ecologically valid, to investigate the relationship between length of exposure and phonological perception, previous instruction should be treated as a variable. To investigate the interaction of [h] with preceding vowels in NSs, more items per vowel and more vowels should be included. More research is required to ascertain to what degree DFA is psychologically real for Germans, and whether RT effects differ between cases in which the listening target is involved in the assimilation rather than adjacent to (but independent of) segments involved in the assimilation.

ACKNOWLEDGEMENTS

This research was supported by a 2015 *Language Learning* Dissertation Grant and by Prof. Dr. Sabine Zerbian at the Universität Stuttgart under the Sonderforschungsbereich-72 Short-Term Research Grant. Additional thanks are due to Isabelle Darcy, Philip LeSourd, Ken de Jong, Susanne Even, Jeff Holliday, Joan-Carles Mora, and the members of the Second Language Psycholinguistics Lab at Indiana University for their consultation and support.

ABOUT THE AUTHOR

John H. G. Scott (scottjh@indiana.edu) received his Ph.D. from the Department of Second Language Studies at Indiana University. His research interests include interaction of phonetic categories and phonological knowledge in L2 perception and implications of phonological

acquisition for other aspects of acquisition. He has taught German for over a decade and seminars on constructed languages and miscommunication.

Contact information:
Department of Second Language Studies
Indiana University
Morrison Hall 216
1165 E. Third Street
Bloomington, IN 47405
Email: scottjh@indiana.edu

REFERENCES

Barriuso, T. A., & Hayes-Harb, R. (2018). High variability phonetic training as a bridge from research to practice. *The CATESOL Journal*, *30*, 177–194. Retrieved from http://www.catesoljournal.org/wp-content/uploads/2018/03/CJ30.1_barriuso.pdf

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*, 607–652. doi: 10.1515/ling.2003.020

Hall, T. A. (2014). Alveolopalatalization in Central German as markedness reduction. *Transactions of the Philological Society, 112*, 143–166. doi: 10.1111/1467-968X.12002

Halle, M., & Stevens, K. N. (1971). A note on laryngeal features. *Quarterly Progress Report, 101*, 198–212. Cambridge, MA: Research Laboratory of Electronics.

Jessen, M. (1998). *Phonetics and phonology of tense and lax obstruents in German*. Amsterdam, Netherlands: John Benjamins.

Lindsey, J. (2013). L2 acquisition of the context-dependent [ç]-[x] alternation in American English learners of German. Unpublished manuscript, Indiana University, Bloomington.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences (Versions 2.8 & 2.9). *Behavior Research Methods, 44*, 314–324. [Computer program]. doi:10.3758/s13428-011-0168-7

Otake, T., Yoneyama, K., Cutler, A., & Van der Lugt, A. (1996). The representation of Japanese moraic nasals. *Journal of the Acoustical Society of America, 100*, 3831–3842. doi: 10.1121/1.417239

Scott, J. H. G. (2019). *Phonemic and phonotactic inference in early interlanguage: Americans learning German fricatives in L2 acquisition* (Unpublished doctoral dissertation). Indiana University: Bloomington.

Segui, J., Frauenfelder, U., & Hallé, P. (2001). Phonotactic constraints shape speech perception: Implications for sublexical and lexical processing. In E. Dupoux (Ed.), *Language, brain, and cognitive development: Essays in honor of Jacques Mehler* (pp. 195–208). Cambridge, MA: MIT Press.

Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics, 39,* 456–466. doi: 10.1016/j.wocn.2010.09.001

Weber, A. (2001a). Help or hindrance: How violation of different assimilation rules affects spoken language processing. *Language and Speech, 44,* 95–118. doi: 10.1177/00238309010440010401

Weber, A. (2001b). *Language-specific listening: The case of phonetic sequences*. Wageningen, The Netherlands: Ponsen & Looijen.

Weber, A. (2002). Assimilation violation and spoken-language processing: A supplementary report. *Language and Speech, 45,* 37–46. doi: 10.1177/00238309020450010201

Wiese, R. (1996). *The phonology of German*. Oxford, UK: Clarendon Press.

---

[i] German NSs self-reported dialects including Swabian, Westphalian, Palatine, Bavarian, (Thuringian-)Franconian, "Hessisch-Platt," Standard German, or none, and all reported knowledge of English. Additional language exposure in this group included French, Spanish, Latin, Swedish, Russian, Italian, Dutch, Hindi, Marathi, Malayalam, and Turkish. I agree with an anonymous reviewer that individuals' different exposure to other L2s may be an important factor for variation. Practical constraints on field work did not allow recruitment of a more uniform German NS group, and individual case studies are beyond the scope of this study.

Thomson, R. I., Foote, J. A. (2019). Pronunciation teaching: Whose ethical domain is it anyways?. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Ames, IA, September 2018 (pp. 226-236). Ames, IA: Iowa State University.

PRESENTATION/POSTER

PRONUNCIATION TEACHING: WHOSE ETHICAL DOMAIN IS IT ANYWAYS?

Ron I. Thomson, Brock University
Jennifer A. Foote, University of Alberta

Several recent articles and book chapters have raised ethical concerns about practices within the field of second language (L2) pronunciation teaching. In this paper, we propose a preliminary set of ethical guidelines for teaching L2 pronunciation, based on a review of related research, and from relevant ethics and standards documents developed by professional associations for North American English Language Teachers (ELTs) (e.g., TESOL and TESL Canada) and Speech Language Pathologists (SLPs) (e.g., ASHA and SAC). We then apply these ethical guidelines to archived data from a survey of 60 ELTs, and 71 SLPs, who offer what they describe as pronunciation instruction or accent modification/accent reduction services. The survey examined instructor qualifications, and teacher knowledge about L2 pronunciation and its teaching. Mixed results indicate that while some ELTs and SLPs appear to adhere to reasonably defined ethical guidelines and standards, many do not. We conclude with recommendations for positive change in this area.

INTRODUCTION

To the extent that ethical standards and codes of conduct should be viewed as a hallmark of a profession's maturity, English language teaching for adult learners remains in its infancy. Not even TESOL International, the largest professional association of English language teachers (ELTs), has an official ethical code of conduct. TESL Canada, a Canadian association of English language teachers, does have a set of ethical guidelines, but it is unknown to what extent they are recognized or taught as a component of TESL Canada recognized teacher training programs. Furthermore, because they are guidelines, and not official policy, there is no enforcement mechanism. Universities are also rarely under any obligation to introduce ethics, and when they do so, it is often in cases where the degree being sought results in licensure by professional associations that have already adopted their own particular code of ethics (e.g. law, medicine, etc.) (Davis, Hildt, & Laas, 2016).

While the absence of a professional and enforceable code of ethics for language teachers of adult learners should, itself, be deeply concerning to the wider language teaching profession, we are particularly interested in ethical practice as it pertains to a subset of language teachers of adults – those who teach second language (L2) pronunciation. This area is of particular interest because relative to teaching of the primary language skills (i.e., reading, writing, listening and speaking), there is far less of an evidence-base to which instructors and learners can turn to determine if pronunciation instruction is warranted or effective (Thomson & Derwing, 2015). The pronunciation sub-domain is also of interest because it is unregulated (Lippi-Green, 2012), and known to often blur the professional boundaries of ELTs, speech language pathologists (SLPs) and entrepreneurs (Derwing & Munro, 2015; Foote, 2018). In addition, consumers of such services comprise a highly vulnerable population (see Thomson, 2014). Many have come to believe that their foreign accent is to blame for communication difficulties, whether this is actually the case or not (Derwing, Fraser, Kang, & Thomson, 2014). While it is true that foreign and even regional first language accents may be the subject of negative evaluation by listeners, resulting discrimination cannot be easily disentangled from discrimination based on other talker characteristics (e.g., race, socio-economic status, etc.). Yet, because foreign accent is so salient, it often becomes the focus of false promises by individuals claiming to be able to quickly eliminate learners' accents and in doing so, improve communication, job prospects and relationship skills (Thomson, 2014).

Previous literature examining the ethics of pronunciation instruction is limited (see Foote, 2018 for a detailed overview), but the literature that does exist expresses similar concerns about a lack of regulation. The purpose of this paper is to propose a preliminary set of ethical guidelines for pronunciation instruction, and then to use existing survey data to determine the extent to which ELTs and SLPs are currently practicing ethical pronunciation instruction. We chose to focus on these two types of practitioners because they provide the overwhelming majority of pronunciation instruction to L2 learners. Further, we are only addressing the teaching of English pronunciation, although the same principles may apply to pronunciation instruction for other L2s.

METHODS

Establishing a set of ethical guidelines for pronunciation instruction

To create a set of ethical guidelines for pronunciation instruction we consulted two data sources: 1) literature that provides evidence-based best practice suggestions for pronunciation instruction and 2) documentation from associations that govern and/or represent ELTs and SLPs. Since TESOL International has no code of ethics, we relied on TESL Canada (2018) to identify ethical statements that we felt were particularly relevant to teaching pronunciation in the North American context. Similarly, we consulted materials from the American Speech-Language Hearing Association (ASHA) (2011) and Speech-Language and Audiology Canada (SAC) (2018), both of which have explicit ethical codes published on their websites. Unlike TESL Canada, ASHA (2011) and SAC (2018) materials make it clear that their members are required to uphold their Codes of Conduct.

Survey data

The survey data used to determine the extent to which ELTs and SLPs apply ethical principles to the teaching of English pronunciation or in the provision of what SLPs often term foreign accent modification (FAM) services were extracted from a much larger study of ELTs and SLPs (see Thomson, 2013 which examined SLP and ELTs background knowledge regarding second language pronunciation and its instruction). While this survey largely followed the format developed by Foote et al. (2011), most of the items were new, and required respondents to agree or disagree with statements taken verbatim from pronunciation and FAM materials found in written texts and on the web. After answering questions about their background qualifications, participants were asked to respond to a series of statements concerning the nature of a foreign accent and instructional strategies and techniques. Although they indicated their agreement using the labels 'strongly agree',

‘agree’, ‘disagree’, ‘strongly disagree’, and ‘unsure’, we have collapsed these into three categories: ‘agree’, ‘disagree’ and ‘unsure’, in order to be more succinct.

Participants were recruited via targeted messages to colleagues, via email lists of relevant interest groups, and through social media forums (e.g., relevant LinkedIn groups). Data used for the purpose of the current study were limited to respondents who self-identified as ELTs or SLPs and who indicated that they taught pronunciation or provided accent modification/reduction services. The ELT group comprised 60 respondents (45 in Canada; 15 in the United States) and 71 SLPs (49 in Canada; 21 in the United States; one with work experience in both). Most respondents were female (80% of ELTs; 94% of SLPs). Most were also native speakers of English (85% of ELTs; 100% of SLPs). Respondents were highly educated. For ELTs, 62% had master’s degrees in ELT, with most of the rest (30%) having related bachelor’s degrees (e.g., TESL, linguistics, etc.). For SLPs, 94% had a master’s degree (the expected credential for licensure), while one had a PhD in the field, and one a bachelor’s. Many respondents (60% of ELTs; 72% of SLPs) reported having taken courses related to pronunciation instruction during their university programs. However, only 19% of ELTs and 3% of SLPs reported taking an entire university course directly related to L2 pronunciation instruction and/or FAM. Many respondents had attended related workshops offered at professional conferences (66% of ELTs and 34% of SLPs).

RESULTS

Our evaluation of peer-reviewed pronunciation teaching and learning literature, and professional guidelines and ethical codes for ELTs and SLPs lead us to propose the following set of eight ethical guidelines that we feel are most applicable to contexts where intensive pronunciation instruction is provided, with relevant citations. These guidelines are not applicable to incidental pronunciation instruction as part of traditional language classes. Further, it is important to note that these eight guidelines are not meant to replace the broader ethical guidelines and codes that exist for these professions. Rather, they are intended to add to and elaborate upon them in ways that are of concern to the domain of pronunciation instruction.

Ethical guidelines for L2 English pronunciation instruction

1. Pronunciation instruction should primarily focus on intelligibility, rather than reduction of accent (Derwing, Fraser, Kang & Thomson, 2014; Foote, 2018; Isaacs & Trofimovich, 2012; Levis, 2005; Levis, 2018; Kang, Thomson, & Moran, 2018).
2. When teaching pronunciation, an L2 accent should be viewed as a natural part of L2 speech development; an L2 accent is not a speech disorder (Derwing & Munro, 2015; Foote, 2018; Thomson, 2014).
3. Individuals offering instruction should not make exaggerated claims about the efficacy of the instruction they offer, or the results of services or products offered (ASHA, 2016; Derwing et al., 2014; Thomson, 2014).
4. Individuals or organizations offering pronunciation instruction should not use fear-based advertising that demonizes an L2 accent. Advertisements should be honest and appropriate (Foote, 2018; SAC, 2016; Thomson, 2014).
5. Pronunciation instruction should not be continued if such instruction is unnecessary or ineffective (ASHA, 2016; SAC, 2016).

6. Individuals offering pronunciation instruction should have specialized training in pronunciation; a degree in TESL or speech-language-pathology may not be sufficient to qualify someone as an expert of pronunciation (Derwing et al. 2014; Foote, 2018; Thomson, 2014).
7. Professionals offering pronunciation instruction should continue to seek professional development and be aware of new research developments in pronunciation research (TESL Canada, 2018).
8. Professionals offering pronunciation instruction should respect the dignity and rights of all persons without prejudice as to race, religious beliefs, sex, gender identity/gender expression, socioeconomic status, sexual orientation, physical characteristics (ASHA, 2016; SAC, 2016; TESL Canada, 2018).

Extent to which survey data from ELTs and SLPs reflects ethical practice

From the existing survey data, we identified items that we feel best reflect examples of adherence to or conflict with our proposed set of ethical guidelines for L2 English pronunciation instruction and report responses to each item in the tables that follow.

Below we provide separate tables to evaluate each guideline. Statements in Table 1 suggest that most ELTs and SLPs hold the appropriate view that the focus of instruction should be on improving intelligibility, not accent. Nevertheless, many SLPs (42%) felt that they were able to eliminate/nearly eliminate a client's accent. Combined with SLPs' greater acceptance of the terms *accent reduction* and *accent modification*, this suggests that SLPs may be more likely than ELTs to offer prospective clients services that are unnecessary, and to place foreign accent in a negative light.

Table 1

Pronunciation instruction should primarily focus on intelligibility, rather than reduction of accent. Percentage agreement with relevant statements

Statement	*Agree*	*Disagree*	*Unsure*
1. The goal of a pronunciation program should be to eliminate, as much as possible, foreign accents.	12% (ELTs) 11% (SLPs)	75% (ELTs) 75% (SLPs)	13% (ELTs) 14% (SLPs)
2. Language teachers are able to eliminate or nearly eliminate a learner's accent.	8% (ELTs) 8% (SLPs)	82% (ELTs) 72% (SLPs)	10% (ELTs) 20% (SLPs)
3. Accent modification/reduction specialists are able to eliminate/nearly eliminate a client's accent.	15% (ELTs) 42% (SLPs)	57% (ELTs) 44% (SLPs)	28% (ELTs) 14% (SLPs)
4. The goal of pronunciation teaching should be to help make students comfortably intelligible to their listeners, even if they still have a strong accent.	87% (ELTs) 82% (SLPs)	2% (ELTs) 10% (SLPs)	12% (ELTs) 8% (SLPs)

5. The goal of pronunciation or accent modification training is not to erase an accent but rather to learn a new accent that will improve communication ability.	75% (ELTs) 80% (SLPs)	7% (ELTs) 7% (SLPs)	18% (ELTs) 13% (SLPs)
6. Someone can have a very strong accent and still be highly intelligible and comprehensible.	82% (ELTs) 72% (SLPs)	15% (ELTs) 20% (SLPs)	3% (ELTs) 8% (SLPs)
7. How comfortable are you with the term 'Accent reduction?'	38% (ELTs) 72% (SLPs)	43% (ELTs) 21% (SLPs)	18% (ELTs) 7% (SLPs)
8. How comfortable are you with the term 'Accent therapy'	13% (ELTs) 20% (SLPs)	70% (ELTs) 66% (SLPs)	17% (ELTs) 14% (SLPs)

While not in the majority, a sizeable percentage of SLPs (24%) explicitly view a foreign accent as analogous to disordered speech (see Table 2). While not as many ELTs (8%) are explicit in this acknowledgment, more ELTs than SLPs tend to implicitly view foreign accent as a pathology. For example, many ELTs feel that pronunciation difficulty is related to muscle weakness or airflow, with far fewer SLPs sharing similar beliefs.

Table 2

When teaching pronunciation, an L2 accent should be viewed as a natural part of L2 speech development; an L2 accent is not a speech disorder. Percentage agreement with relevant statements

Statement	*Agree*	*Disagree*	*Unsure*
1. A foreign accent is not unlike other communication disorders.	8% (ELTs) 24% (SLPs)	62% (ELTs) 68% (SLPs)	30% (ELTs) 8% (SLPs)
2. Errors in pronunciation result from not having speech muscles that are properly toned for English sounds. Articulation exercises are critical.	73% (ELTs) 28% (SLPs)	10% (ELTs) 49% (SLPs)	17% (ELTs) 23% (SLPs)
3. Increase your range of motion by moving your chin from side to side and up and down.	27% (ELTs) 20% (SLPs)	25% (ELTs) 63% (SLPs)	48% (ELTs) 17% (SLPs)
4. Instruments placed in your mouth that position the tongue correctly can be used to correctly pronounce words with an American accent.	12% (ELTs) 14% (SLPs)	43% (ELTs) 63% (SLPs)	45% (ELTs) 23% (SLPs)
5. Improper air-flow is a common cause of a foreign accent.	40% (ELTs) 20% (SLPs)	18% (ELTs) 61% (SLPs)	42% (ELTs) 20% (SLPs)

Many ELTs and SLPs agree with exaggerated claims about the efficacy of pronunciation instruction (see Table 3), but SLPs agree with such beliefs more frequently. Only a minority disagree with these exaggerated claims, but many more are unsure, suggesting that they may be susceptible to adopting such beliefs.

Table 3

Individuals offering instruction should not make exaggerated claims about the efficacy of the instruction they offer, or the results of services or products offered. Percentage agreement with relevant statements

Statement	*Agree*	*Disagree*	*Unsure*
1. For a significant and permanent reduction in your accent, you need to see a specialist.	23% (ELTs) 58% (SLPs)	47% (ELTs) 17% (SLPs)	30% (ELTs) 25% (SLPs)
2. Pronunciation/accent instructors can help clients learn to turn on or off many of their accented sounds whenever the need arises.	45% (ELTs) 62% (SLPs)	15% (ELTs) 13% (SLPs)	40% (ELTs) 25% (SLPs)
3. In private classes, students can experience major success in as little as 2 hours.	30% (ELTs) 39% (SLPs)	30% (ELTs) 25% (SLPs)	40% (ELTs) 35% (SLPs)
4. Internet coaching can make a dramatic change in people's accent.	32% (ELTs) 31% (SLPs)	3% (ELTs) 6% (SLPs)	65% (ELTs) 63% (SLPs)
5. Students need only practice for five minutes every day to experience good results in a month.	27% (ELTs) 11% (SLPs)	30% (ELTs) 59% (SLPs)	43% (ELTs) 30% (SLPs)

Both ELTs and SLPs seem to agree that using fear-mongering advertising that paints accent in a bad light are inappropriate (e.g., accents limit personal, educational and career success) (see the first two statements in Table 4). Neither, however, seem to object to advertising that suggests *not* speaking with an accent will lead to some competitive advantage, although ELTs approve of this reverse fear-mongering in smaller numbers (see the latter two statements in Table 4).

Table 4

Individuals or organizations offering pronunciation instruction should not use fear-based advertising that demonizes an L2 accent. Advertisements should be honest and appropriate. Percentage agreement with relevant statements

Statement	*Agree*	*Disagree*	*Unsure*
1. An accent does not mean you don't know how to speak a language, but it may limit you at work and at home.	82% (ELTs) 85% (SLPs)	12% (ELTs) 11% (SLPs)	7% (ELTs) 4% (SLPs)
2. A foreign accent will limit educational and career choices.	23% (ELTs) 48% (SLPs)	47% (ELTs) 27% (SLPs)	30% (ELTs) 25% (SLPs)

3. Accent modification training can provide you with a distinct competitive advantage.	68% (ELTs) 82% (SLPs)	5% (ELTs) 1% (SLPs)	27% (ELTs) 17% (SLPs)
4. Employees who have completed accent modification training are more confident, effective communicators who enjoy greater job satisfaction.	47% (ELTs) 73% (SLPs)	3% (ELTs) 1% (SLPs)	50% (ELTs) 25% (SLPs)

The statements in Table 5 reflect what we believe are dubious beliefs about the necessity of ongoing pronunciation instruction and support techniques that have no theoretical or empirical support. Many ELTs and SLPs believe that teaching pronunciation does not result in permanent change and so would offer ongoing instruction. Treatments for which the efficacy is unproven is surely not something that should be continued. Only a small number support the use of unproven designer techniques. Many, however, are unsure about whether such techniques are effective and so may be more likely to use such techniques if suggested by others.

Table 5

Pronunciation instruction should not be continued if such instruction is unnecessary or ineffective. Percentage agreement with relevant statements

Statement	*Agree*	*Disagree*	*Unsure*
1. Teaching pronunciation does not usually result in permanent changes; ongoing treatment is necessary.	42% (ELTs) 37% (SLPs)	28% (ELTs) 42% (SLPs)	30% (ELTs) 21% (SLPs)
2. Instruments placed in your mouth that position the tongue correctly can be used to correctly pronounce words with an American accent.	12% (ELTs) 14% (SLPs)	43% (ELTs) 63% (SLPs)	45% (ELTs) 23% (SLPs)
3. Practicing speaking with a pencil in your mouth will help you direct your attention to your articulators.	22% (ELTs) 6% (SLPs)	33% (ELTs) 72% (SLPs)	45% (ELTs) 23% (SLPs)
4. Final consonants are very, very, aggressive in America, the final consonant needs to be deleted to not sound angry. For example, you should say "wha" not "what".	10% (ELTs) 6% (SLPs)	75% (ELTs) 92% (SLPs)	15% (ELTs) 3% (SLPs)

Regarding the qualifications necessary to provide pronunciation instruction (see Table 6), both ELTs and SLPs have a relatively high confidence in their ability. Nevertheless, some (approximately 20%) offer pronunciation training despite not feeling that they are qualified to do so. There is less agreement, by both groups, as to whether specialized training in pronunciation instruction should be a prerequisite to offering it.

Table 6

Individuals offering pronunciation instruction should have specialized training in pronunciation; a degree in TESL or speech-language-pathology may not be sufficient to qualify someone as an expert of pronunciation. Percentage agreement with relevant statements

Statement	*Agree*	*Disagree*	*Unsure*
1. I am completely comfortable teaching segmentals	80% (ELTs) 94% (SLPs)	8% (ELTs) 4% (SLPs)	12% (ELTs) 1% (SLPs)
2. I am completely comfortable teaching suprasegmentals	85% (ELTs) 86% (SLPs)	12% (ELTs) 4% (SLPs)	3% (ELTs) 10% (SLPs)
3. Pronunciation instruction should only be offered by instructors who have taken courses specific to pronunciation AND/OR accent modification	43% (ELTs) 68% (SLPs)	48% (ELTs) 18% (SLPs)	14% (ELTs) 14% (SLPs)
4. Do you believe you are qualified to offer pronunciation instruction?	82% (ELTs) 78% (SLPs)	3% (ELTs) 6% (SLPs)	15% (ELTs) 16% (SLPs)

The lack of consensus as to what constitutes adequate preparation in pronunciation instruction is also reflected in Table 7. While the majority of both groups wish that they had more training in how to teach pronunciation, a relatively large number of ELTs (33%) and SLPs (48%) do not.

Table 7

Professionals offering pronunciation instruction should continue to seek professional development and be aware of new research developments in pronunciation research. Percentage agreement with relevant statements

Statement	*Agree*	*Disagree*	*Unsure*
1. I wish I had more training in teaching pronunciation	67% (ELTs) 52% (SLPs)	15% (ELTs) 34% (SLPs)	18% (ELTs) 15% (SLPs)

Finally, there is some evidence that some ELTs and SLPs hold views that show a lack of knowledge about their clients. While most rightly agree that the physiology of particular groups (i.e., jaw shape) has nothing to do with a foreign accent, many are unsure or believe that it does. Also connected to the concept of accent and race, a large number of ELTs and SLPs believe that only native speakers should be teaching pronunciation.

Table 8

Professionals offering pronunciation instruction should respect the dignity and rights of all persons without prejudice as to race, religious beliefs, sex, gender identity/gender expression, socioeconomic status, sexual orientation, physical characteristics. Percentage agreement with relevant statements

Statement	*Agree*	*Disagree*	*Unsure*
1. Some ethnic groups have jaw shapes that make learning English pronunciation difficult	12% (ELTs) 4% (SLPs)	55% (ELTs) 77% (SLPs)	33% (ELTs) 18% (SLPs)
2. Only native speakers should teach pronunciation.	20% (ELTs) 38% (SLPs)	62% (ELTs) 35% (SLPs)	18% (ELTs) 27% (SLPs)

DISCUSSION

As noted in the introduction, professional ethics might reasonably be considered a mark of a profession's maturity. It seems sensible, then, to expect an ethical code of conduct for those providing pronunciation instruction, since the recipients of such instruction represent a population that is particularly susceptible to exploitation in this area. This is particularly true for learners who may themselves believe that eliminating their foreign accent is a realistic goal.

The title of this article asks whose ethical domain is pronunciation instruction? We cannot say with confidence that this domain currently belongs to either ELTs or SLPs. Having proposed what we believe are a reasonable and important set of ethical guidelines for those offering pronunciation instruction, our analysis of ELTs' and SLPs' beliefs and practices reveals that neither group appears to be sufficiently ethical in their practice.

In some cases, ELTs and SLPs might learn from each other. SLPs are more likely than ELTs to treat accent negatively, which may in part be due to their motivation as business people (see Thomson, 2014). ELTs are more likely to hold erroneous beliefs about the underlying source of a foreign accent, believing it to be due to motor-speech difficulty, while SLPs are more likely to understand that this is not the case.

Another common theme in our analysis of ELTs' and SLPs' beliefs and practices related to ethical pronunciation instruction is the high degree of uncertainty many respondents feel in evaluating beliefs that those in their field hold, and practices that many in their field use. This should be taken as evidence that they need more training specific to L2 pronunciation instruction. Yet, most feel that they are qualified, and many do not have a desire for further training to develop their skills. This disconnect, between their lack of certainty in what they do, but confidence in their educational background, needs to be addressed.

IMPLICATIONS

We hope that the ethical guidelines proposed here will provide the foundation for further work in this area and in particular, promote ethical practice within this domain. Ultimately, given the population that pronunciation instructors serve and evidence that L2 learners are, in many cases, not receiving ethical instruction, it is imperative that formal ethical guidelines be established. Ethical guidelines are the domain of professional associations for ELTs and SLPs, who need to work with content area experts to enforce ethical practice for this subset of the populations that they serve.

ACKNOWLEDGEMENTS

We are thankful to the ELTs and SLPs who responded to our survey.

ABOUT THE AUTHORS

Ron Thomson is a Professor of Applied Linguistics at Brock University, in Niagara, Canada. His research interests include L2 oral fluency and pronunciation development, professional ethics, high variability phonetic training, and computer assisted pronunciation teaching.

Contact information:
Department of Applied Linguistics, Brock University
500 Glenridge Ave
St. Catharines, ON, L2S 3A1
Canada
email: rthomson@brocku.ca
tel: 905-688-5550 ext 5842
fax: 905-688-2360

Jennifer Foote is an Assistant Professor in Applied Linguistics in the Faculty of Extension, at the University of Alberta, in Edmonton, Canada. Her research interests include beliefs and practices around pronunciation instruction, ethical practice, and pronunciation instructional interventions.

REFERENCES

American Speech-Language Hearing Association (ASHA). (2016). *Code of ethics.* Retrieved from www.asha.org

Davis, M., Hildt, E., & Laas, K. (2016). Twenty-five years of ethics across the curriculum: An assessment. *Teaching Ethics, 16*(1), 55-74. doi: 10.5840/tej201633028

Derwing, T. M., Fraser, H., Kang, O. & Thomson, R. I. (2014). L2 accent and ethics: Issues that merit attention. In A. Mahboob & L. Barratt (Eds), *Englishes in multilingual contexts*. Berlin: Springer.

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

Foote, J. A. (2018). Ethics and the business of pronunciation instruction. In O. Kang, R. I. Thomson, & J. M. Murphy (Eds.), *The Routledge Handbook of Contemporary English Pronunciation.* (pp. 284-297). London: Routledge.

Foote, J. A., Holtby, A. K, & Derwing, T. M. (2011). Survey of pronunciation teaching in adult ESL programs in Canada, 2010. *TESL Canada Journal, 29*(1), 1-22.

Kang, O., Thomson, R. I., & Moran, M. (2018). Which features of accent affect understanding? Exploring the intelligibility threshold of diverse accent varieties. *Applied Linguistics*. Advanced online publication.

Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly, 39*(3), 369-377.

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

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

Speech-Language and Audiology Canada (SAC). (2016). *Code of Ethics*. Retrieved from www.sac-aoc.ca

TESL Canada. (2018). *Ethical Guidelines for English as a Second Language Professionals.* Retrieved from www.tesl.ca

Thomson, R. I. (2013). ESL teachers' beliefs and practices in pronunciation teaching: Confidently right or confidently wrong? In J. Levis & K. LeVelle (Eds.), *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference.* Aug. 2012. (pp. 224-233). Ames, IA: Iowa State University.

Thomson, R. (2014). Accent reduction and pronunciation instruction are the same thing. In L. Grant (Ed.), *Pronunciation myths: Applying second language research to classroom teaching*, 160-187. Ann Arbor, MI: University of Michigan Press.

Thomson, R. I., & Derwing, T. M. (2015). The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics, 36*(3): 326-344.

Ueda, R., & Hashimoto, K. (2019). Perceptual training in a classroom setting: Phonemic category formation by Japanese EFL learners. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Ames, IA, September 2018 (pp. 237-249). Ames, IA: Iowa State University.

PRESENTATION/POSTER

PERCEPTUAL TRAINING IN A CLASSROOM SETTING: PHONEMIC CATEGORY FORMATION BY JAPANESE EFL LEARNERS

Ruri Ueda, Osaka Kyoiku University
Ken-ichi Hashimoto, Osaka Kyoiku University

Perceptual training targeting L2 phonemes has been reported as effective for both L2 learners' perception and production learning even without articulation practice. Considering the situation in EFL countries, especially Japan, where most English teaching and learning occur in classrooms with limited time, perceptual training can be an easy-to-conduct, effective method for L2 sound acquisition. Many of the studies reporting its positive effects, however, examined lab-based training, and only a few studies have tested the effects of perceptual training in a classroom setting. Therefore, to examine the applicability of perceptual training in the classroom, in the present study a ten-minute perceptual training targeting English /b/−/v/, /l/−/r/, and /s/−/θ/ was conducted in English courses at a university in Japan for six weeks. The results showed that students' scores on both the perception and production of /b/−/v/ and /s/−/θ/ significantly improved. However, the learning was not generalized to new word stimuli. For the /l/−/r/ contrast, neither their perception nor production performance changed after the training. Some possible reasons for smaller training effects than reported in many studies are discussed with reference to a lack of sufficient input and the way feedback was provided.

INTRODUCTION

For successful communication, it is necessary to perceive and produce the sounds of the language(s) used in communication. When it comes to L2 communication, it is widely agreed that learners' L1 has an influence on their L2 pronunciation performance (e.g., Tsukada, Birdsong, Bialystok, Mack, Sung, & Flege, 2005). Perception and production of L2 sounds absent from the learners' L1 sound system are said to be difficult because they do not have L2 phonetic representations or proper phonetic categories (e.g., Cutler, 2012; Flege, 1992). Several influential models have been proposed to explain the degree of difficulty in mastering each L2 phoneme according to the learner's L1, such as the Perceptual Assimilation Model (PAM; Best, 1995) and Flege's Speech Learning Model (SLM; 1995). According to PAM, the degree of difficulty perceiving each phoneme depends on how its phonemic contrasts are assimilated to learners' L1 phonemic categories. On the other hand, SLM explains that the greater the perceived dissimilarity of an L2 sound from the closest sound of the learner's L1, the more likely the acquisition of the L2 sound is. SLM also hypothesizes that perception and production share underlying representations, suggesting that improving perception skills by constructing phonetic representations can guide production learning as well.

Given the importance of skills to deal with L2 sounds in communication and the difficulty of mastering them, L2 educators have been concerned about effective ways of constructing L2 phonetic representations or proper L2 phonetic categories, building on the speech perception models described previously. One possible way is perceptual training, and many studies have reported its positive effects on learners' skills in perceiving and producing L2 sounds (e.g., Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Lambacher, Martens, Kakehi, Marasinghe, & Molholt, 2005; Thomson, 2011). A typical format of perceptual training is an identification task with two-alternative forced choice (2AFC) format. In the task targeting the English /l/−/r/ contrast, for instance, participants first hear a stimulus like "lead." The minimal pair "lead" and "read" is then presented visually, and participants choose the word they think they heard. Immediate feedback is provided following the participant's choice. This type of perceptual training is considered effective for constructing L2 phonetic representations or modifying L2 phonetic categories because learners have exposure to extensive L2 sound input focusing on the target phonemes. More striking is that training could potentially lead to improvement of both L2 perception and production skills. For instance, Bradlow and her colleagues have extensively tested the influences of perceptual training on L2 learners' productions of the target sounds. In Bradlow et al. (1997), the participants were Japanese college students, and the target phonemic contrast was /l/ and /r/. The participants received perceptual training with the 2AFC format for 15 to 22.5 hours. The results showed that both their perception and production skills for the L2-English /l/−/r/ contrast improved. Also, the learning was generalized to non-familiar talkers and new words. A subsequent study reported that the learning effects persisted even three months after training (Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999).

These findings suggest that perceptual training is effective, and that exposure to L2 sounds is important to improve L2 perception and production skills. However, L2 learners in EFL countries, such as Japan, have very limited opportunities to receive L2 sound input outside of the classroom. Given the importance of input in L2 acquisition (e.g., Thomson, 2011), it is ideal, and even essential, to provide sufficient sound input in class so that learners can acquire the skills to both perceive and produce sounds in the target language. We assume that perceptual training is an effective way to achieve this goal.

Most studies of the effects of perceptual training are lab-based, and only a few have reported on perceptual training in a classroom setting. In Hamada and Tsushima (2001), Japanese college students had three-week perceptual training on seven phonemic contrasts. The pre and posttest comparison found training effects on both perception and production skills. However, this study assigned out-of-class training sessions as homework as well as in-class training, which made it difficult to examine the effects of classroom-based training alone. Considering the possibility of applying perceptual training in the classroom, it is essential to examine whether the training effects observed in lab-based studies are also found in the classroom. Therefore, the present study examined the effects of classroom-based perceptual training for L1-Japanese learners on the pronunciation of L2-English consonants. The following research questions were addressed:

1. Does classroom-based perceptual training improve both the perception and production of L2-English phonemes?
2. Does classroom-based perceptual training generalize to accurate perception and production of new stimuli that did not appear during training?

METHODOLOGY

Participants

The study was conducted in two weekly English communication classes at a Japanese university for L1-Japanese college students. One class served as a treatment group (n = 24) and the other served as a control (n = 25). Only the data from the participants who attended every session during the study period were used in the analysis, leaving 13 students in each group. The overall English proficiency of the control group was slightly higher than that of the treatment group, which requires us to interpret results with caution.

Procedure

The study had a pre and posttest design. In week 1, both groups had a pretest session (see details below). From weeks 2 to 7, the treatment group had a weekly 10-minute perceptual training at the beginning of class, and the control group had regular English conversation practice instead. In week 8, both groups took a posttest session. In weeks 9 to 14, the content of the two groups was flipped in consideration of research ethics.

Training

The training took the format of a 2AFC task with the High Variability Training technique (Lively, Logan, & Pisoni, 1993), which consisted of stimuli spoken by four talkers (two female, two male). The target contrasts were /b/−/v/, /l/−/r/, and /s/−/θ/, which L1-Japanese speakers have difficulty in both perceiving and producing (Lambacher, 1999). Words used in the training were base−vase, berry−very, best−vest, bought−vote, lane−rain, late−rate, lead−read, lock−rock, seem−theme, sick−thick, sing−thing, and sum−thumb. The participants had 16 trials for each contrast each day of the training session, which means they had 48 trials in total per day.

The major difference of the present perceptual training from many reported studies was being conducted in class. The training video was projected on the screen at the front of the classroom so that all the students could do the task at the same time. The video contained a set of sound stimuli, a visual probe of the minimal pair, and immediate feedback for each training item, the last to facilitate the participants checking their answer immediately on their own. Each trial started with a beep, then the participants heard a word, after which they had to mark their answer on the answer sheet before a chime sounded. The correct answer was then zoomed and colored in red as feedback. The next trial then started with another beep.

Pre and posttests

The pre and posttests comprised two tasks, a 2AFC task to assess learners' perceptual skills and a word-list reading task for production skills. The former task was conducted as in the training session except that no feedback was provided in the tests. The pretest had 96 trials and the posttest had the same 96 trials and 24 additional trials to test learning generalization. For analysis, the posttest stimuli were divided into four conditions. Stimuli A were the same training stimuli spoken by two of the four speakers used in the training session; Stimuli B were the same training words spoken by non-familiar talkers; Stimuli C were new words spoken by two of the talkers used in the training session; and Stimuli D were new words spoken by non-familiar talkers. Conditions B, C, and D were established to examine the generalizability of the perceptual training.

In the word-list reading task, the participants were recorded reading aloud randomly-listed words from the posttest identification task by themselves. Later, the speech data were judged by three L1-English speakers with a 2AFC task format.

RESULTS

The perception skill results

Table 1 shows the perception scores for the training words spoken by familiar talkers in the pre and posttests.

Table 1

The identification task results of the training words spoken by familiar speakers

		Treatment group		Control group	
Contrast		Pre	Post	Pre	Post
/b/−/v/	*Mean*	10.85	13.38	12.46	12.85
	S.D.	2.51	1.61	3.13	1.99
/l/−/r/	*Mean*	8.62	8.85	9.31	9.69
	S.D.	2.75	2.48	2.25	2.78
/s/−/θ/	*Mean*	11.31	13.92	12.69	12.92
	S.D.	2.25	1.32	2.63	1.89
Totals	*Mean*	30.77	36.15	34.46	35.46
	S.D.	5.64	4.62	4.93	4.05

Overall, the treatment group significantly improved, while the control group did not. A series of two-way mixed ANOVAs with Time as a within-participant factor and Group as between-participant were conducted on the participants' identification scores of the /b/−/v/, /l/−/r/, and /s/−/θ/ contrasts. The alpha level was set at .017 to avoid Type I Error in multiple ANOVAs. For the /b/−/v/ contrast, the main effect of Time was observed [$F(1, 24) = 9.92$, $p < .004$, partial $\eta 2 = .293$], while Group effect was not significant [$F(1, 24) = 0.44$, $p = .512$, n.s., partial $\eta 2 = .018$]. The Time × Group interaction approached significance [$F(1, 24) = 5.39$, $p < .029$, partial $\eta 2 = .183$]. Post hoc analyses using a Holm's Sequentially Rejective Bonferroni Procedure revealed a significant difference in the scores between the pre- and post-identification tasks in the treatment group [$F(1, 12) = 23.25$, $p < .001$, partial $\eta 2 = .660$], but not in the control group [$F(1, 12) = 0.25$, $p < .624$, n.s., partial $\eta 2 = .021$]. As for the /l/−/r/ contrast, none of the main effects nor their interaction was significant ($Fs < 1$). The results for the /s/−/θ/ contrast were almost identical to those for the /b/−/v/ contrast. The main effect for Time was significant [$F(1, 24) = 14.80$, $p < .001$, partial $\eta 2 = .381$], while the Group effect was not significant [$F(1, 24) = 0.07$, $p = .794$, n.s., partial $\eta 2 = .003$]. The Time × Group interaction was significant [$F(1, 24) = 10.39$, $p < .004$, partial $\eta 2 = .302$]. Post hoc analyses revealed a significant difference in only the treatment group [$F(1, 12) = 20.10$, $p < .001$, partial $\eta 2 = .626$], not in the control group [$F(1, 12) = 0.26$, $p < .621$, n.s., partial $\eta 2 = .021$].

Table 2 shows the posttest perception scores for the training words spoken by non-familiar talkers.

Table 2

The identification task results of the training words spoken by non-familiar speakers

		Treatment group	Control group
/b/−/v/	*Mean*	10.00	9.23
	S.D.	2.88	1.48
/l/−/r/	*Mean*	8.69	10.15
	S.D.	2.97	2.90
/s/−/θ/	*Mean*	14.38	12.69
	S.D.	1.08	1.81
Totals	*Mean*	33.08	32.08
	S.D.	5.43	4.97

Overall, the scores of the two groups were almost the same between the pre and posttests. An advantage for the treatment group was found in the /s/−/θ/ contrast, but not in the /b/−/v/ and /l/−/r/ contrasts.

Table 3 shows the posttest scores when the participants listened to new words.

Table 3

The identification task results of the new words spoken by familiar and non-familiar speakers

Contrast		Treatment group		Control group	
		Familiar	Non-familiar	Familiar	Non-familiar
/b/−/v/	*Mean*	2.62	2.46	2.92	2.92
	S.D.	0.74	0.63	0.73	0.73
/l/−/r/	*Mean*	1.92	2.00	2.38	3.00
	S.D.	1.14	1.18	1.08	1.04
/s/−/θ/	*Mean*	3.54	3.15	3.23	3.38
	S.D.	0.63	0.66	0.58	0.62
Totals	*Mean*	8.08	7.62	8.54	9.31
	S.D.	1.21	1.50	1.74	1.68

When the talkers were familiar, the control group's scores were better than the treatment group's overall. However, a series of two-way Group × Familiarity ANOVAs performed on each of the contrasts with an adjusted alpha level of .017 showed no significant main effects or interaction (Fs < 4.38). The trend was almost the same when the talkers were not familiar. This time, the control group was consistently better, but again, only numerically.

The production skill results

Table 4 shows the production results for the training words, which suggest that the treatment group showed a larger improvement than the control group.

Table 4

The production results for the training words

Contrast		Treatment group		Control group	
		Pre	Post	Pre	Post
/b/−/v/	*Mean*	12.92	16.15	16.08	17.23
	S.D.	2.10	4.95	3.64	4.64
/l/−/r/	*Mean*	13.85	14.69	13.31	14.31
	S.D.	3.72	5.42	4.33	4.64
/s/−/θ/	*Mean*	13.62	18.62	16.00	17.92
	S.D.	2.02	3.64	2.83	4.07
Totals	*Mean*	40.38	49.46	45.38	49.46
	S.D.	5.33	10.38	6.16	7.53

The scores for each contrast showed particularly large improvements for the /b/−/v/ and /s/−/θ/ contrasts in the treatment group, as suggested by two-way Time × Group mixed ANOVAs (alpha adjusted at .017 again). For the /b/−/v/ contrast, the main effect of Time was significant [$F(1, 24) = 7.54$, $p = .011$, partial $\eta 2 = .239$], while the Group effect was not [$F(1, 24) = 2.47$, $p = .129$, n.s., partial $\eta 2 = .093$]. The Time × Group interaction was not significant [$F(1, 24) = 1.69$, $p = .206$, n.s., partial $\eta 2 = .066$], though larger improvement was found in the treatment group. For the /l/−/r/ contrast, neither of the main effects nor their interaction was significant ($Fs < 1.81$). For the /s/−/θ/ contrast, the main effect of Time was significant [$F(1, 24) = 23.74$, $p < .001$, partial $\eta 2 = .497$]. Neither the Group effect [$F(1, 24) = 0.65$, $p = .429$, n.s., partial $\eta 2 = .026$] nor the Time × Group interaction [$F(1, 24) = 4.69$, $p = .041$, partial $\eta 2 = .164$] was significant. As the interaction effect approached significance, post hoc analyses were performed, revealing a significant difference in the average scores between the scores for the first and second recordings in both the treatment [$F(1, 12) = 30.95$, $p < .001$, partial $\eta 2 = .721$] and control groups [$F(1, 12) = 6.11$, $p =.021$, partial $\eta 2 = .203$].

Finally, Table 5 shows the production results for the new words.

Table 5

The production results for the new words

Contrast		Treatment group Pre	Treatment group Post	Control group Pre	Control group Post
/b/−/v/	*Mean*	4.00	4.23	4.77	4.46
	S.D.	1.41	1.42	1.36	1.61
/l/−/r/	*Mean*	3.62	3.00	3.00	3.31
	S.D.	1.04	1.96	1.63	1.63
/s/−/θ/	*Mean*	3.54	4.38	3.92	4.31
	S.D.	2.03	1.56	1.38	1.11
Totals	*Mean*	11.15	11.62	11.69	12.08
	S.D.	3.29	3.25	2.21	2.69

The same series of two-way mixed ANOVAs (Time × Group) with adjusted alpha of .017 were conducted. For the /b/−/v/ contrast, there were no significant main or interaction effects (Fs < 1.16), suggesting that neither group showed observable improvement. For the /l/−/r/ contrast, the main effects and their interaction were all non-significant (Fs < 3.18). Finally, for the /s/−/θ/ contrast, the treatment group showed improvement, while the control group did not, though the improvement did not reach significance (Fs < 2.60).

Correlation analyses

Another way to test the effectiveness of the perceptual training is to examine the correlations of the perception and production scores. Larger correlation coefficients should be observed for the scores for contrasts where the perceptual training had a positive effect. Table 6 shows the results of the correlation analysis of the perception and production scores for the pre and posttests. Curiously, we observed larger correlation coefficients on the posttest for all the contrasts in the treatment group, including the /l/−/r/ contrast that showed little training effect.

Table 6

The correlation coefficients of the perception and production scores

Contrast	Treatment group pre	Treatment group post	Control group pre	Control group post
/b/−/v/	-.19	.42	.47	.48
/l/−/r/	.12	.71	.39	.53
/s/−/θ/	.27	.60	.34	.09
Totals	.26	.69	.52	.57

To consider this further, we examined how the scores of individual participants changed. Figure 1 is a correlation plot of the overall scores for the treatment group, which mostly showed constant improvements in perception and production performance. This suggests that their phonetic representations of the target phonemes stabilized or their L2 phonetic categories were modified during training.

A similar trend was observed for the /s/–/θ/ contrast, as shown in Figure 2. Most of the participants improved in both perception and production, suggesting more stable representations or modified phonetic categories.

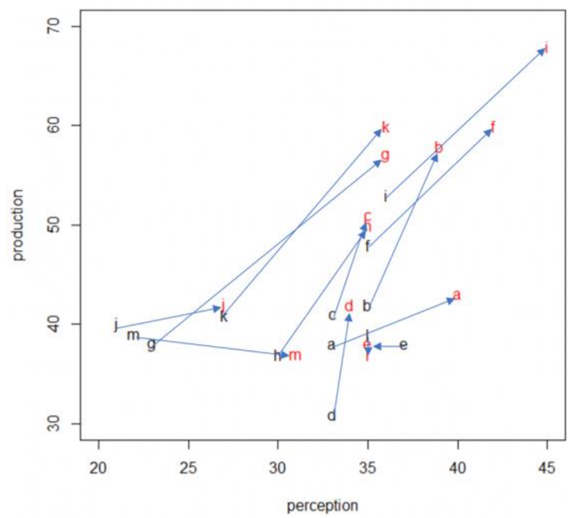


Figure 1. Transitions of overall perception and production scores by participants between the pre and posttests.

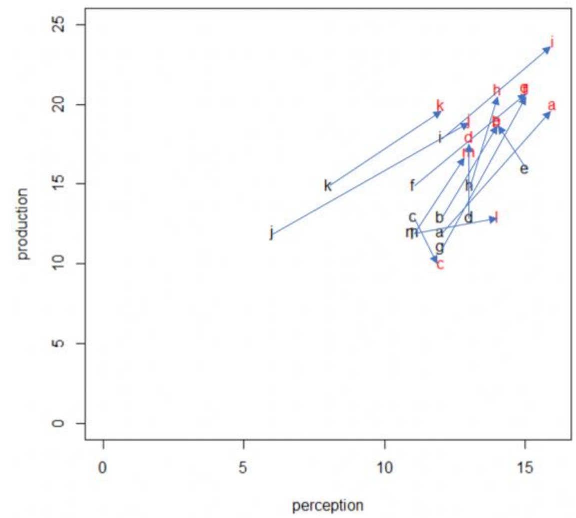


Figure 2. Transitions of /s/–/θ/ perception and production scores by participants between the pre and posttests.

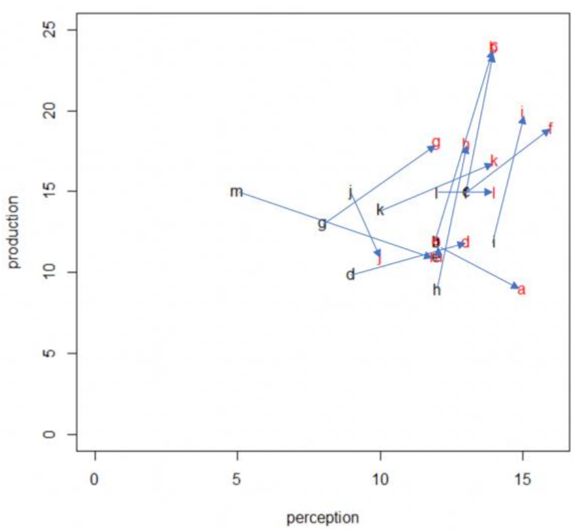


Figure 3. Transitions of /b/–/v/ perception and production scores by participants between the pre and posttests.

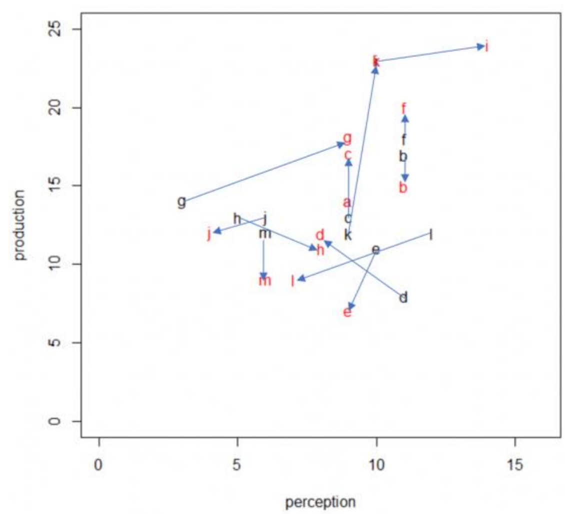


Figure 4. Transitions of /l/–/r/ perception and production scores by participants between the pre and posttest.

As for the /b/−/v/ contrast shown in Figure 3, again, we see constant improvement in perception, but the production scores changed more randomly. Some showed a dramatic improvement, while others were worse on the posttest. These somewhat mixed changes may have led to the weaker correlation on the posttest.

Finally, Figure 4 is a plot for the /l/−/r/ contrast, which showed no clear patterns. Although some participants improved in both perception and production, others showed a totally opposite trend, getting lower scores for both perception and production on the posttest. The posttest scores clustered around the regression line, which may have caused higher correlation coefficients on the posttest.

DISCUSSION

To sum up the results with reference to the research questions mentioned above, the answer to Research Question 1 is partially affirmative, with evidence that the present perceptual training improved the participants' perception and production of the familiar /b/−/v/ and /s/−/θ/, but not /l/−/r/ sounds. As for Research Question 2, however, the answer was negative, as the learning effect did not generalize to untrained stimuli except for the training words of the /s/−/θ/ contrast spoken by non-familiar talkers.

The training conducted in the present study was not effective in changing learners' /l/−/r/ perceptual and productive performances. One factor is the differences in how the three contrasts are assimilated to learners' L1 categories based on the PAM (Best, 1995). The English /l/−/r/ contrast is categorized as "single-category assimilation" by Japanese L1 learners, which means that both sounds are assimilated to the same L1 category to the same extent. On the other hand, the English /b/−/v/ and /s/−/θ/ contrasts are categorized as showing "category-goodness difference," that is, "Both sounds are assimilated to the same L1 category, but one is a far better match to it than the other" (Cutler, 2012, p. 306). According to the PAM, "single-category assimilation" is more difficult to learn than "category-goodness difference," making /l/−/r/ more difficult to acquire than /v/ and /θ/ for the present participants.

In addition, differences in the manner of articulation might also have affected the learning difficulties for each contrast. According to a meta-analysis by Sakai and Moorman (2018), perceptual training was more effective on obstruents than sonorants. They suggested that because obstruent sounds are articulated more saliently, learners can perceive the differences in sound more easily, which facilitates their learning new phonemes. In the present study, the English /b/−/v/ and /s/−/θ/ contrasts involve obstruents, while the /l/−/r/ contrast sonorants. In particular, discriminating English /l/−/r/ requires detecting formant differences. However, L1-Japanese speakers have difficulty in utilizing formant information to discriminate English /l/−/r/. For example, although F3 frequency plays an important role in discriminating English /l/−/r/ sounds, L1-Japanese speakers tend to rely on F2 frequency, which is insufficiently reliable in /l/−/r/ discrimination (Iverson, Hazan, & Bannister, 2005). Therefore, catching the differences underlying the /l/−/r/ contrast should have been more difficult for the present participants.

Regarding the learning generalization, the training effects were not generalized to new stimuli in either perception or production. In studies reporting learning generalization, the participants had 15–22.5 hours of training for each pair of target phonemes, while a recent study of Qian, Chukharev-Hudilainen, and Levis (2018) reported that learning was not generalized to new words when the participants had only 10- to 100-minute training for 12 contrasts. Because the training conducted in the present study lasted only an hour for six phonemes, the lack of sufficient input might be the primary reason why the training effect failed to generalize.

Another possible reason for the lack of learning generalization is the way feedback was provided. Most of the studies that reported positive effects of perceptual training were lab-based. However, the training conducted in the present study was classroom-based. A major difference between the two is how learners receive feedback. In lab-based training, learners receive individualized feedback, while in classroom-based training, answers are presented on the screen item by item, which the learners need to check by themselves. In such a situation, inattentive learners can easily miss the correct answer and do not notice whether or not they made a mistake. Previous studies showed that corrective feedback facilitates improvement of L2 speech perception (Lee & Lyster, 2015) and production (Lyster, Saito, & Sato, 2013), since feedback gives L2 learners opportunities to modify their knowledge. Failure to utilize feedback information effectively might be another source of the decreased learning effects in the present study.

CONCLUSION AND FUTURE STUDIES

The present study examined the effects of perceptual training of L2-English phonetic contrasts in a classroom setting for L1-Japanese college students in Japan. The training led to improvements in the /b/−/v/ and /s/−/θ/ contrasts but not the /l/−/r/ contrast in both perception and production. Moreover, while the training effects were generalized to the training words spoken by non-familiar talkers for the /s/−/θ/ contrast, none of the contrasts were generalized to new words.

There are two issues that need to be addressed in the future studies. First, as mentioned in the discussion, the amount of input is critical. Therefore, we would like to conduct perceptual training with more input to see if learners' /l/−/r/ performances improve. The other critical issue concerns feedback. In the present study, learners were simply shown the correct answer for each item on the screen. In a future study, learners will be asked to mark their answers by themselves to elicit greater attention to the correctness of their answers to see whether this leads to more effective learning. Eventually, in future studies we would like to examine the proposed models of L2 phonetic acquisition and the relations between L2 sound perception and production.

ABOUT THE AUTHORS

Ruri Ueda, M.Ed. (ruri.u10@gmail.com), is a research student in the Department of English Education at Osaka Kyoiku University. She has researched the effects of classroom-based perceptual training and flipped classrooms. Also, she presented a study of differences in L2 learners' pronunciation performances by degree of task spontaneity at the 9th PSLLT Conference. She is currently interested in the relationship between L2 perception and production at the segmental level.

Ken-ichi Hashimoto, Ph.D. (kenhashi@cc.osaka-kyoiku.ac.jp), is an associate professor in the Department of English Education at Osaka Kyoiku University. His research interests lie in the psycholinguistics of second language acquisition, and he obtained his PhD degree from the University of Queensland, Australia, with a dissertation on the effects of L1 syntactic properties on L2 relative clause processing. Since his PhD study, he has been interested in various factors affecting L2 sentence comprehension processes, including L1 influences and interactions between lower-level (e.g., phonetic and lexical) and higher-level (e.g., syntactic) processes.

Contact information for both authors:
Osaka Kyoiku University
4-698-1, Asahigaoka, Kashiwara,
Osaka, 582-8582, Japan
+81-72-978-3525

REFERENCES

Best, C. T. (1995). A direct realist perspective on cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research* (pp. 167–200). Timonium, MD: York Press.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception and Psychophysics, 61*, 977–985. doi:10.3758/BF03206911

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. I. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America, 101*(4), 2299–2310. doi:10.1121/1.418276

Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words*. Cambridge: MIT Press.

Flege, J. E. (1992). The intelligibility of English vowels spoken by British and Dutch talkers. In R. D. Kent (Ed.), *Intelligibility in speech disorders: Theory, measurement, and management* (pp. 157–232). Amsterdam: John Benjamins. doi:10.1075/sspcl.1.06fle

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). Timonium, MD: York Press.

Hamada, M., & Tsushima, T. (2001). Relation between improvements of perception and production ability during a university English speech training course. In J. White (Ed.), *Proceedings of the Fourth Conference on Foreign Language Education and Technology* (pp. 579–582). Kobe: Japan Association for Language Education and Technology.

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. doi:10.1121/1.2062307

Lambacher, S. (1999). A CALL tool for improving second language acquisition of English consonants by Japanese learners. *Computer Assisted Language Learning,12*(2), 137–156. doi:10.1076/call.12.2.137.5722

Lambacher, S. G., Martens, W. L., Kakehi, K., Marasinghe, C. A., & Molholt, G. (2005). The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Applied Psycholinguistics, 26*(2), 227–247. doi:10.1017/S0142716405050150

Lee, A. H., & Lyster, R. (2016). The effects of corrective feedback on instructed L2 speech perception. *Studies in Second Language Acquisition, 38*(1), 35–64. doi:10.1017/S0272263115000194

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America, 94*(3), 1242–1255. doi:10.1121/1.408177

Lyster, R., Saito, K., & Sato, M. (2013). Oral corrective feedback in second language classrooms. *Language Teaching, 46*(1), 1–40. doi:10.1017/S0261444812000365

Qian, M., Chukharev-Hudilainen, E., & Levis, J. (2018). A system for adaptive high-variability segmental perceptual training: Implementation, effectiveness, transfer. *Language Learning & Technology, 22*(1), 69–96. doi:10125/44582

Sakai, M., & Moorman, C. (2018). Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Applied Psycholinguistics, 39*(1), 187–224. doi:10.1017/S0142716417000418

Thomson, R. I. (2011). Computer assisted pronunciation training: Targeting second language vowels: Perception improves pronunciation. *CALICO Journal*, *28*(3), 744–765. doi:10.11139/cj.28.3.744-765

Tsukada, K., Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Flege, J. (2005). A developmental study of English vowel production and perception by native Korean adults and children. *Journal of Phonetics, 33*(3), 263–290. doi:10.1016/j.wocn.2004.10.002

Guskaroska, A. (2019). The potential of ASR for facilitating vowel pronunciation practice for Macedonian learners. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Ames, IA, September 2018 (pp. 250-260). Ames, IA: Iowa State University.

PRESENTATION/POSTER

THE POTENTIAL OF ASR FOR FACILITATING VOWEL PRONUNCIATION PRACTICE FOR MACEDONIAN LEARNERS

Agata Guskaroska, Iowa State University

The purpose of this study is to examine Automated Speech Recognition (ASR) software and its potential for facilitating vowel pronunciation practice for Macedonian English as a Foreign Language (EFL) learners. A list of 12 sentences including minimal pairs of the contrasts /i/-/ɪ/, /æ/-/ɛ/, /u/-/ʊ/, and /ɑ/-/ʌ/ was recorded by 10 Macedonian learners, aged 18-19 and two American English native speakers in order to test the reliability of ASR. The speech samples were turned into text using ASR and the results of the written output were compared between native speakers and non-native speakers. Results demonstrated that the program was accurate in transcribing most of the vowel sounds for native speech. ASR written output was less accurate for non-native speech and was most likely indicating learners' mispronunciations of vowels by transcribing them inaccurately. The results suggest that ASR may be promising for individual vowel practice but future research may involve words in isolation to avoid the system's flaws in making assumptions based on context.

INTRODUCTION

The process of second language acquisition requires development of several aspects of the second language. One of the areas which is usually neglected by instructors, possibly due to lack of time, desire, or training, is pronunciation (Huensch, 2018). Learners often express a desire to work on their pronunciation (LeVelle & Levis, 2014; McCrocklin & Link, 2016), but pronunciation is a skill that requires feedback and is difficult to acquire autonomously (McCrocklin, 2016). In this digital era, researchers are exploring technology with the aim of finding appropriate tools that can assist L2 pronunciation improvement by providing feedback to learners (Levis & Suvorov, 2014; Wallace, 2016).

In that regard, several studies have explored the effectiveness and the potential of ASR (such as *Dragon NaturallySpeaking, Google web speech, and Siri*) and its ability to assist learners by providing feedback with the text-to-speech written output (Derwing, Munro, & Carbonaro, 2000; McCrocklin, 2016; Mroz, 2018). Levis & Suvorov (2014) define ASR as "an independent, machine-based process of decoding and transcribing oral speech" (p. 1) which turns the speech signal into text. Findings from previous studies (e.g., Coniam 1998; Derwing, Munro, & Carbonaro, 2000; Eskenazi, 1999) have mostly indicated that ASR was not fully developed to provide reliable feedback to the learners. Nonetheless, these researchers agreed that if ASR were to improve in the future, it could provide a wide range of possibilities for language learning.

Nonetheless, recent studies found generally positive results towards use of ASR for pronunciation practice (Liakin, Cardoso, & Liakina, 2014; McCrocklin, 2016; Mroz, 2018). These studies pointed towards establishment of learner autonomy and progress. ASR has tremendous potential in applied linguistics and learners appreciate its use (Levis & Suvorov, 2014). It looks promising for pronunciation self-access work and can provide a safe environment for learners. While past research is mostly in favor of ASR, researchers pointed out that the software's accuracy needs further exploration. In that regard, more research is needed to examine whether ASR has improved throughout time. Therefore, this study will examine Macedonian learners' use of ASR as a way to test the system's accuracy.

Contrast between the Macedonian and the English vowel system

The phonetic system of the Macedonian standard language includes five vowels: /i/, /e/, /a/, /o/ and /u/. In English, there are arguably around 12 vowels and eight diphthongs (Dodd & Mills, 1996). In this paper we focus on the American accent variety. Because there are many more vowels in English than in Macedonian, almost every English vowel presents a potential pronunciation problem for Macedonian learners and may be classified as non-existent in the Macedonian language (Kirkova-Naskova, 2012). Even /ɛ/ which is acoustically closest, is pronounced differently depending on phonetic context and dialect region the Macedonian learner belongs to. Figure 1 depicts a comparison between Macedonian and English vowels diagrams.

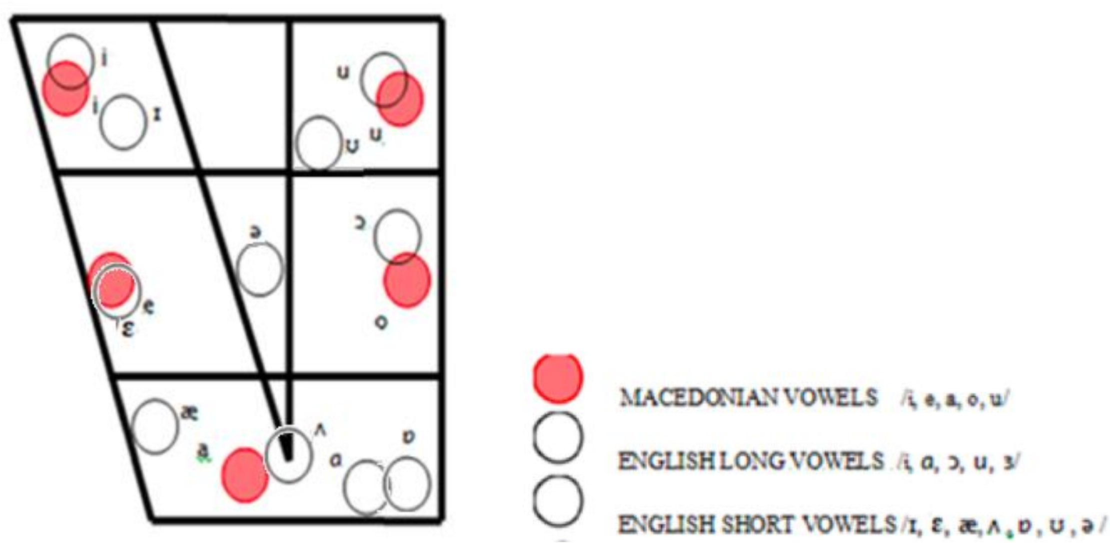


Figure 1. Macedonian and English vowel diagrams (adapted from Krikova-Naskova, 2012).

Based on the comparison above, the first selected vowel pair that might be problematic for Macedonian EFL learning is /i/-/ɪ/. This contrast is very frequent and difficult for these learners because, in Macedonian, there is only one sound which is somewhere in between these two sounds (Kirkova-Naskova, 2009). The minimal pair /æ/-/ɛ/ is also an important contrast because /æ/ does not exist in Macedonian while /e/ is the closest with the English /ɛ/. The /u/-/ʊ/ contrast is similar to /i/-/ɪ/ in terms of difficulty of perception by Macedonian learners. Even though the contrast between these two sounds is not frequent in English and has a low functional load (Munro & Derwing, 2006), Macedonian learners rarely hear the difference between these sounds. Finally, none of the sounds /ɑ/ and /ʌ/ exist in Macedonian and learners very often substitute them with the Macedonian sounds /o/ or /a/, respectively. Hence, this study includes the following vowel contrasts: /i/-/ɪ/, /æ/-/ɛ/, /u/-/ʊ/, and /ɑ/-/ʌ/.

The study

Inspired by the lack of research in this field, as well as lack of tools to assist the EFL/ESL classrooms, this study investigates the potential of an ASR tool, *Apple's Enhanced Dictation Feature*, for providing corrective feedback to Macedonian learners. *Apple's Enhanced Dictation* is available in OS X Mavericks v10.9 or later. This program is free, easily available, and user-friendly and for those reasons it was selected for this study. This study explores ASR's accuracy by comparing ASR's recognition of native and non-native speech. The exploration of the written output of NSs will show the accuracy for these speakers which can help in the exploration of ASR's potential to facilitate vowel pronunciation practice for Macedonian learners.

Research questions

The following questions guide this study:

1. How accurate is the ASR program Enhanced Dictation in recognizing and transcribing native English speakers' production of vowel contrasts?
2. How accurate is the ASR program in recognizing and transcribing Macedonian L2 learners' production of vowel contrasts?

METHODOLOGY

Participants

The participants who took part in this research were Macedonian EFL learners, aged 18-19, who provided speech samples of their English for evaluation. 10 Macedonian native speakers, seven male and three female (level B2, n=4; and C1, n=6) according to the Common European Framework of Reference for Languages, were recorded. None of the participants had lived for any period in an English-speaking country.

The control group included two male, native speakers (NS) of American English, graduate students familiar with pronunciation, aged 28-29. The NSs speech was recorded to provide a standard for comparison in order to evaluate the program and its ability to turn words into text.

Materials and procedure

The materials consisted of 12 sentences containing minimal pairs that were the same parts of speech (e.g. "The patient wanted to leave."; "The patient wanted to live."). See the Appendix for the minimal pairs. These minimal pairs were deliberately chosen to be the same parts of speech to avoid the program's assumptions of certain words based on their position in the sentence. Based on the comparison between the English vowel systems, the selected sounds were included (/i/-/ɪ/; /æ/-/ɛ/; /u/-/ʊ/; and /ɑ/-/ʌ/). All the vowel contrasts had three instances containing the vowel, for example, for the sound /i/, the words *leave, sleep* and *sheep* were chosen. The selected words consisted of simple vocabulary that is introduced at low levels of EFL classes and hence known to the students to avoid problems in pronunciation due to lack of knowledge of the word meaning. ASR used for turning the voice into speech for this study was *Apple's Enhanced Dictation*.

The participants were encouraged to first read the sentences once to themselves quietly and then to record themselves while they read the sentences aloud at a normal pace. Participants recorded their speech with a voice recording application on their phone (iPhone or Android) and sent the recordings via email. The speech samples were played to ASR, the speech was turned into text and saved as text in a Microsoft word document.

Data analysis

The words were manually evaluated for accuracy, then counted separately per vowel and total and turned into percentages. In order to evaluate the program's accuracy for recognizing native and non-native speech, the data were separately analyzed and summarized in tables. After that, a comparison of the results was made between NS and NNS's written output to identify differences and similarities.

FINDINGS

Recognition of native speakers of English

To answer the first research question, the ASR's written output of NSs speech was analyzed. During the analysis, the focus was only on the targeted minimal pair words, not the entire sentence they were embedded in. The sentences only served to provide context because the purpose of this study was to focus on vowel contrasts.

Table 1

Number and percent of accurate and inaccurate recognized lexical items by the ASR program (Native English speakers)

No. of participants	No. of lexical items per speaker	Total No. of lexical items	Accurate	Inaccurate
2	24	48	42 (87.5%)	6 (12.5%)

Overall, the program did not provide 100% accuracy when it comes to vowel recognizing and turning voice into speech for NSs, in the context used. The program examined showed 87.5% total accuracy, which is close to what several similar studies found. For instance, Derwing et al. (2000) found 90% accuracy and Ashwell and Elam (2017) found 89.4%. Even though the total accuracy of the program in this study is 87.5%, analyzing each sound recognition individually can provide us with a clearer picture of the tool's capabilities. Table 2 provides a closer look into each sound and identifies the sounds which the program failed to recognize.

Table 2

Number and percent of accurate and inaccurate recognized vowel contrasts by the ASR program (Native English speakers)

Lexical items		Leave Sleep Sheep	Live Slip Ship	Pan Laughed Man	Pen Left Men	Luke Wooed Boot	Look Would Book	Cop Dock Shot	Cup Duck Shut
Vowels		i	ɪ	æ	ɛ	u	ʊ	ɑ	ʌ
Accurate items									
	No.	6	6	6	4	6	6	6	2
	%	100%	100%	100%	66%	100%	100%	100%	33%
Inaccurate items									
	No.				2				4
	%				33%				66%
Total									
	No.	**6**	**6**	**6**	**6**	**6**	**6**	**6**	**6**
	%	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**

Interestingly, almost all the NSs' vowels were transcribed 100% correctly with the exception of the vowels /ɛ/ and /ʌ/. Regarding the vowel /ɛ/, the only lexical item that ASR did not recognize was the word *men*. The sentences used for this commonly mistaken vowel contrasts were: *I saw the man with the yellow coat* and *I saw the men with the yellow coat*. ASR failed to recognize the plural form of this word in all the instances, which made the recognition of /ɛ/ 66% accurate. If it is not just a challenging pair, perhaps the system relies on context to assist in word recognition. In other words, the system may suppose the singular form of the word and thus transcribes the word as *man* in both sentences. Future research could explore the accuracy of the program by isolating the words and not providing any context. On the other hand, the program transcribed all the other /æ/-/ɛ/ words correctly, thus proved accurate in this study by 100% for recognizing /æ/ and 66% for /ɛ/ sound.

Another vowel that was unrecognized was /ʌ/. Only 33% of the words containing the vowel /ʌ/ were recognized and transcribed correctly. The issue with the recognition of these sentences may be an indicator of the program's assumption based on context. For example, the sentence *I sat on the duck* was contrasted to the sentence *I sat on the dock*. One possible explanation is that the word *dock* may likely appear more often in this type of context and hence the program may have transcribed the word incorrectly merely making an assumption based on frequency. Regardless of the reasons, these findings show that the ASR program did not appear to be highly reliable for the sound /ʌ/ used in this context. In this study the overall accuracy of the system's recognition of NS vowels appears to be high, nonetheless, it might be important to consider vocabulary and context selection in order to avoid the system's possible limitations.

Recognition of non-native L1 Macedonian ESL speech

To answer the second research question, I calculated the number and percentage of recognized vowels in the system's written output of the targeted minimal pairs of NNS. The overall score of accuracy for NNS was 71%, as seen in Table 3. These findings align with Derwing et al. (2000) study where they found that the software was 71-73% accurate for nonnative speech for Cantonese and Spanish L1 learners, while Ashwell and Elam (2017) found 65.7% for Japanese and a few Chinese speakers.

Table 3

Number and percent of accurate and inaccurate recognized lexical items by ASR (Macedonian ESL learners)

No. of participants	**No. of lexical items per speaker**	**Total No. of lexical items**	**Accurate**	**Inaccurate**
10	24	240	170 (71%)	70 (29%)

Nonetheless, the overall results present the systems' accuracy in general, and do not give a clear picture about each targeted vowel. Table 4 summarizes the findings for each sound separately to get a better overview of the situation.

Table 4

Number and percent of accurate and inaccurate recognized vowel contrasts by ASR (Macedonian EFL learners)

Lexical items	Leave Sleep Sheep	Live Slip ship	Pan Laughed Man	Pen Left Men	Luke Wooed boot	Look Would book	Cop Dock shot	Cup Duck shut
Vowels	i	ɪ	æ	ɛ	u	ʊ	ɑ	ʌ
Accurate items								
No.	22	22	17	20	16	29	29	15
%	73%	73%	57%	67%	53%	97%	97%	50%
Inaccurate items								
No.	8	8	13	10	14	1	1	15
%	27%	27%	43%	33%	47%	3%	3%	50%
Total								
No.	**30**	**30**	**30**	**30**	**30**	**30**	**30**	**30**
%	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**

When looking at individual sounds produced by the L2 learners, we can note that no sound was recognized with 100% accuracy, although the sounds /ʊ/ and /ɑ/ are close, both at 97% recognition. The lowest percentage of accuracy was with /ʌ/ with 50% accuracy. However, when comparing to the NSs, the system was not considered reliable regarding the sounds /ʌ/ by transcribing only 33% of NS words correctly. This may suggest that the overall system struggled to recognize this sound. Other sounds with low recognition were /u/ with 53% and /æ/ with 57%. Both /i/ and /ɪ/ showed 73% accuracy. As discussed earlier, this pair was expected to be difficult for Macedonian learners. However, the pair is also very frequent which might have resulted in better pronunciation then other sounds. On the other hand, when it comes to the sounds /æ/ and /u/, ASR demonstrated 100%

accuracy for native speech and only 43% and 47% accuracy, respectively. These findings suggest that these Macedonian learners may have issues with distinguishing the production of most of the vowel contrasts, considering that ASR transcribed NS accurately.

DISCUSSION

To be useful for vowel pronunciation practice for L2 learners, ASR should first recognize native speech with high accuracy. The overall score of recognition was lower than expected with 87.5% accuracy (see Table 1). The non-native speech was transcribed less accurately (71% as shown in Table 3). Is this an indicator that the program cannot transcribe a non-native speech, or is it an indicator that the program gives good feedback to the learners because it writes what it ‘hears’? This percentage might be interpreted as the general ability of ASR to indicate intelligibility and, as Wallace (2016) points out, to suggest the words which were unclear. Even though previous studies criticized the ability of ASR to recognize non-native speech (Coniam, 1999; Derwing et al., 2000), more recently ASR tools have been improving and several recent studies are in favor of the program for L2 pronunciation practice (Liakin et al., 2014; McCrocklin, 2016; Mroz, 2018; Wallace, 2016).

The overall results may align with previous studies, however, the overview of individual sounds shows that almost all the vowels were transcribed 100% correctly for NS with the exception of the vowels /ɛ/ and /ʌ/. Regardless of whether the selected pairs might have been challenging or the program might have ‘assumed’ words out of context, with the exception of these two vowels, ASR showed 100% accuracy for the rest of the vowels for NSs. These findings may be indicators that in future studies, vowel pronunciation practice should be tested by using individual isolated words, instead of sentences, to eliminate the possibility of the influence of context.

On the other hand, ASR did not show the same level of accuracy for identifying individual vowels for Macedonian learners. Was the program identifying vowel mispronunciations? While it cannot be claimed that these errors were due to mispronunciation, as it may be due to other reasons, closer analysis of the output showed that many of the errors seem closely connected to problems with mispronunciation. Kirkova-Naskova (2010) also points out that the most challenging minimal pairs for Macedonian learners is /æ/-/e/, also identifying /u/-/ʊ/,/ɪ/-/i/ and /ʌ/-/a/ as common foreign markers in Macedonian-accented speech. In this study, ASR seemed to be identifying specific vowels that were likely mispronounced by these speakers and present the most common foreign markers of their speech. Hence, it could be argued that the program appeared to be providing feedback to the learners’ mispronunciations and might be considered useful for vowel pronunciation practice for Macedonian learners. In order to confirm this hypothesis, future studies should include NSs’ judgments of the non-native speech.

Previous studies on ASR pointed out that it can be beneficial to students in various ways, such as creating a safe environment for self-practice, saving time, self-monitoring (Wallace, 2016), fostering learner autonomy, supplementing course work (McCrocklin, 2015), and raising students’ awareness (Mroz, 2018). Mroz (2018) found that learners are mostly satisfied with their ASR experience, emphasizing that the written output was a good feedback for them as it provided visual representation of their words. All these benefits make ASR an interesting field that needs further exploration.

This exploratory study for Macedonian learners of English for vowel pronunciation practice showed that, besides exploring the overall accuracy scores, examining the way individual sounds are turned into text can also be valuable and should also be explored when evaluating ASR's accuracy. The findings suggest that ASR was most likely indicating learners' mispronunciations of vowels by transcribing the words inaccurately. Hence, this study may provide evidence that ASR has promising potential for L2 learners vowel pronunciation practice and should be explored further in the future.

CONCLUSION

This study explored the accuracy of an ASR tool, *Apple's Enhanced Dictation*, and its possibility to provide corrective feedback for vowel pronunciation practice in an EFL context. Even though the enhanced dictation feature is limited to Macintosh users, the results suggest that ASR may have great potential for providing corrective feedback to EFL learners for a select set of vowel contrast. Even though the overall accuracy score for NSs was not as high as desired, the program was accurate in this study for recognition of individual vowel sounds for American native speech. The only sound for which the program demonstrated flaws was the sound /ʌ/ (only 33% correct). In terms of Macedonian EFL speech, the ASR written output was less accurate and it was most likely indicating learners' mispronunciation of vowels by transcribing the words inaccurately.

For future studies, words containing the target vowel sounds can be used in isolation to avoid possible influence of the sentence context when the program turns speech into text. Furthermore, to confirm the usefulness of the program, future studies may include native human raters in order to make a comparison between the program's feedback and human judgment. Finally, ASR may be recommended for individual vowel practice for Macedonian EFL classroom use, but further research is required to confirm these findings.

ACKNOWLEDGMENTS

I would like to thank Dr. John Levis for his guidance and support. In addition, I would like to thank Dr. Shannon McCrocklin, Dr. Lara Wallace and Erin Todey, for their valuable feedback and observations on this topic.

ABOUT THE AUTHOR

Agata Guskaroska is a Fulbright scholar and a MA student in TESL/Applied Linguistics at Iowa State University. She taught ESL courses for seven years and American literature at FON University. Her major interests include CALL, pronunciation, and SLA.
Email: agatag@iastate.edu

REFERENCES

Ashwell, T., & Elam, J.R. (2017). How accurately can the Google web speech API recognize and transcribe Japanese L2 English learners' oral production? *JALT CALL Journal, 13*(1), 59-76.

Coniam, D. (1999). Voice recognition software accuracy with second language speakers of English. *System, 27,* 49-64.

Derwing, T. M., Munro, M. J., & Carbonaro, M. (2000). Does popular recognition software work with ESL speech? *TESOL Quarterly*, *34,* 592-603. https://doi.org/10.2307/3587748

Dodd, S., & Mills, J. (1996). *Phonetics and phonology.* CORE, University of Exeter Press. core.ac.uk/display/19510000.

Eskenazi, M. (1999). Using automatic speech processing for foreign language pronunciation tutoring: Some issues and a prototype. *Language Learning & Technology*, *2*(2), 62–76. http://dx.doi.org/10125/25043

Huensch, A. (2018). Pronunciation in foreign language classrooms: Instructors' training, classroom practices, and beliefs. *Language Teaching Research*. 1–20. DOI: 10.1177/1362168818767182

Kirkova-Naskova, A. (2012). Interlanguage phonology: comparison between the English and the Macedonian vowel systems. *Annual Symposium of the Faculty of Philology 'Blaze Koneski'* pp. 141-152. Skopje: 'University St. Cyril and Methodius'.

Kirkova-Naskova, A. (2010). Native Speaker Perceptions of Accented Speech: The English Pronunciation of Macedonian EFL Learners. *Research in Language, 8*, 1-21.

Kirkova-Naskova, A. (2009). *Markers of Foreign Accent in Macedonian-accented English*. Unpublished MA Thesis. Skopje: Faculty of Philology.

Levelle, K., & Levis, J. (2014). Understanding the impact of social factors on L2 pronunciation: Insights from learners. In J. Levis, & A. Moyer (Eds.), *Social dynamics in second language accent* (pp. 97-118). Boston: DeGruyter.

Levis, J., & Suvorov, R. (2014). Automated speech recognition. In C. Chapelle (Ed.), *The encyclopedia of applied linguistics*. Retrieved from *http://onlinelibrary.wiley.com/store/10.1002/9781405198431.wbeal0066/asset/wbeal0066.pdf?v=1&t=htq1z7hp&s=139a3d9f48261a7218270113d3833da39a187e74*

Liakin, D., Cardoso, W., & Liakina, N. (2014). Learning L2 pronunciation with a mobile speech recognizer: French /y/. *CALICO Journal*, *32*(1), 1-25.

McCrocklin, S. (2015). Automatic speech recognition: Making it work for your pronunciation class. In J. Levis, R. Mohammed, M. Qian, & Z. Zhou (Eds.), *Proceedings of the 6th Pronunciation in Second Language Learning and Teaching Conference* (ISSN 2380-9566), Santa Barbara, CA (pp. 126-133). Ames, IA: Iowa State University.

McCrocklin, S. M. (2016). Pronunciation learner autonomy: The potential of Automatic Speech Recognition. *System*, *57*, 25–42.

McCrocklin, S., & Link, S. (2016). Accent, Identity, and a Fear of Loss? ESL Students' Perspectives. *Canadian Modern Language Review, 72*(1), 122-148.

Mroz, A. (2018). Seeing how people hear you: French learners experiencing intelligibility through automatic speech recognition. *Foreign Language Annals,* 1-21. https://doi.org/10.1111/flan.12348

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

Wallace, L. (2016). Using Google web speech as a springboard for identifying personal pronunciation problems. 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. 180-186). Ames, IA: Iowa State University

APPENDIX.

Minimal pair sentences

/u/-/ʊ/

1. **Look/Luke** Look! There's a rabbit over there. Luke! There's a rabbit over there.
2. **Full/Fool** What's the meaning of the word 'full'? What's the meaning of the word 'fool'?
3. **Boot/Book** I lost my boot. I lost my book.

/i/-/ɪ/

4. **Leave/Live** The patient wanted to leave. The patient wanted to live.
5. **Sleep/Slip** Did you sleep on the ice? Did you slip on the ice?
6. **Sheep/Ship** Where's my sheep?/ Where's my ship?

/æ/-/e/

7. **Men/Man** I saw the man with the yellow coat. I saw the men with the yellow coat
8. **Pen/Pan** Can you please give me the pen? Can you please give me the pan?
9. **Left/Laughed** I told her a joke and she left/ I told her a joke and she laughed.

/ʌ/-/ɑ/

10. **Cup/Cop** I don't like that cup. I don't like that cop.
11. **Duck/Dock** He sat on the duck. He sat on the dock.
12. **Shut/Shot** The door was shut. The door was shot.

Guskaroska, A., & Taylor, J. (2019). A corpus-analysis of gendered items in pop and country music from the 90s to now. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching conference*, ISSN 2380-9566, Iowa State University, September 2018 (pp. 261-275). Ames, IA: Iowa State University.

PRESENTATION/POSTER

A CORPUS-ANALYSIS OF GENDERED ITEMS IN POP AND COUNTRY MUSIC FROM THE 90s TO NOW

Agata Guskaroska, Iowa State University
Joshua Taylor, Iowa State University

This study examines how gender has been portrayed in Pop and Country music across the last 30 years. Using a corpus-based approach, we examined the lyrics of 80 songs across the Pop and Country music genres from the 1990s and now. The results suggest that use of language has changed over time and genres and the scope of these nouns is wider than their traditional definitions. Our findings indicate that the word *baby* is the predominant word used to refer to men and women in both genres, and that mostly this refers to women either as a lover or an individual that is attractive or desirable. In addition, Pop music sexually objectifies women frequently while country music typically refers to women as beautiful or in a way that expresses romantic worship or desire.

INTRODUCTION

Different music genres involve different uses of language which provides an interesting area for exploration of various linguistic features, such as vocabulary, grammar, social, and cultural representations of groups of people. Because of this, music lyrics present an interesting ground for linguistic research. One way to analyze the variation between these aspects of language is through corpus-based methods, which are a practical and useful way to examine language use. By using corpora, a large amount of data can be analyzed to explore and compare artists' creative use of language. For instance, researchers have explored lyrics from various genres including Hip-Hop (Daniels, 2014; Shin, 2016), Pop (Motschenbacher, 2016), and Country music (Shin, 2016).

Song lyrics reflect the social and cultural context of a society and present an interesting area for language exploration. Language and music are brought together in songs (Barwick, Birch, & Evans, 2007). The meanings of words can very often go beyond the traditional dictionary definition. In that regard, an interesting trend to explore in song lyrics is the way males and females are portrayed in different music genres. A few studies have explored gendered lemmas (a base word and its inflection, e.g. *girl*, *girls*), such as Pearce (2008) who used a corpus-based approach to examine how the lemmas *man* and *woman* were used in the British National Corpus, a 100 million word collections and samples of written and spoken language, and Baker (2013) who focused on a critical approach of using corpora in sociolinguistics.

Motschenbacher (2016) observed that while skepticism towards using lyrics for language exploration existed, they are increasingly being analyzed. He argues that one reason for this skepticism is the doubt of authenticity of this type of language because lyrics are carefully crafted, adapted, and written for a wide audience. While it is true that lyrics are not the most authentic use of language, the language can spread worldwide and impact the social, cultural, and linguistic habits of generations. A few lyrics corpora have been built so far, including a corpus of Hip-hop lyrics at Carnegie Mellon University (Friginal & Hardy, 2018) and the *Hip-Hop Word Count*, a searchable database containing 40,000 Hip-Hop songs built from 1979 and still growing (Hopkins, 2011).

Different music genres promote different values and perspectives on males' and females' roles in society. One such study which investigated this is Shin (2016), who explored gendered items in Hip-Hop and country music. She found that in Hip-Hop females are often sexually objectified, associated with beauty, and described as dependent. On the other hand, country music usually portrayed female as romantically worshiped and the lemmas *girl* and *woman* often referred to family members. She found that both females and males are still stereotypically presented in song lyrics, especially in Hip-Hop.

While existing studies have shown the sociolinguistic importance of analyzing lyrics, the research of lyrics is limited and there is a need for further exploration. Inspired by Shin's (2016) study, our paper examines how gender has been portrayed in Pop and Country music across the last 30 years. The purpose of our work is to reveal more insightful perspectives and attitudes toward gender, as well as interesting trends of language use in two music genres. The following two research questions guide our study:

- **RQ 1**. What do the four gender lemmas (*girl*, *woman*, *boy*, & *man*, and the neutral-gender lemma *baby*) primarily refer to in Pop and Country music, and how have they changed across time and genres?
- **RQ 2.** How is gender portrayed in Pop and Country music, and how have these portrayals changed across time and genres?

METHODS

Using a mixed methods approach, we analyzed our corpus first with AntConc (Anthony, 2017), and then conducted a qualitative analysis of the pragmatic function of each lemma within its context. We then identified common trends between and within our data sets. We examined the lyrics of 80 songs across Pop and Country music genres from the 1990s and now (20 from each decade for both genres, totaling 80) by analyzing five previously selected lemmas: *girl*, *woman*, *boy*, *man*, and *baby*. Initially, only the four gender lemmas (*boy*, *man*, *girl* and *woman*) were included; however, we discovered that both males and females are much more commonly referred to as *baby* in popular songs and therefore we included it in the analysis. Even though the primary notion of the noun is to refer to "a very young child", the informal definition of *baby* is: "A lover or spouse (often as a form of address)" (Oxford, 2018). Related findings from Motschenbacher's (2016) analysis of Eurovision song lyrics show the noun *baby* is the 17th most common collocate of love. Hence, the neutral lemma *baby* was included in the research to discover its wider scope and reference.

Corpus design and quantitative analysis

In order to obtain a sample of the most popular music, we searched for lists of the top hits in each time period. We used Billboard (2018) which provided complete lists of (1) the top 20 billboard pop hits between 2010-2017, (2) the top 20 billboard pop hits between 1990-1999, and (3) the top

20 country billboard hits between 2010-2017. Billboard did not have the same information for 90's country songs, so this data was found from Buzz Feed (Burton, 2013). We then extracted the lyrics and converted them into .txt files. The corpus data is shown in Table 1.

Table 1

Number of song texts and word counts across genres and time periods

	Pop Music 90s	Contemporary Pop Music	Country Music 90s	Contemporary Country Music	TOTAL
Texts	20	20	20	20	80
Word Count	7412	8250	5140	6378	27180

Each of the sub-corpora was analyzed by using AntConc 3.5.6 developed by Anthony (2017). The selected word lemmas were identified using regular expressions. We retrieved the results from AntConc and saved them as .txt files for further qualitative analysis (see Table 2 below). One example of the search for the lemma *girl* is illustrated in Figure 1.

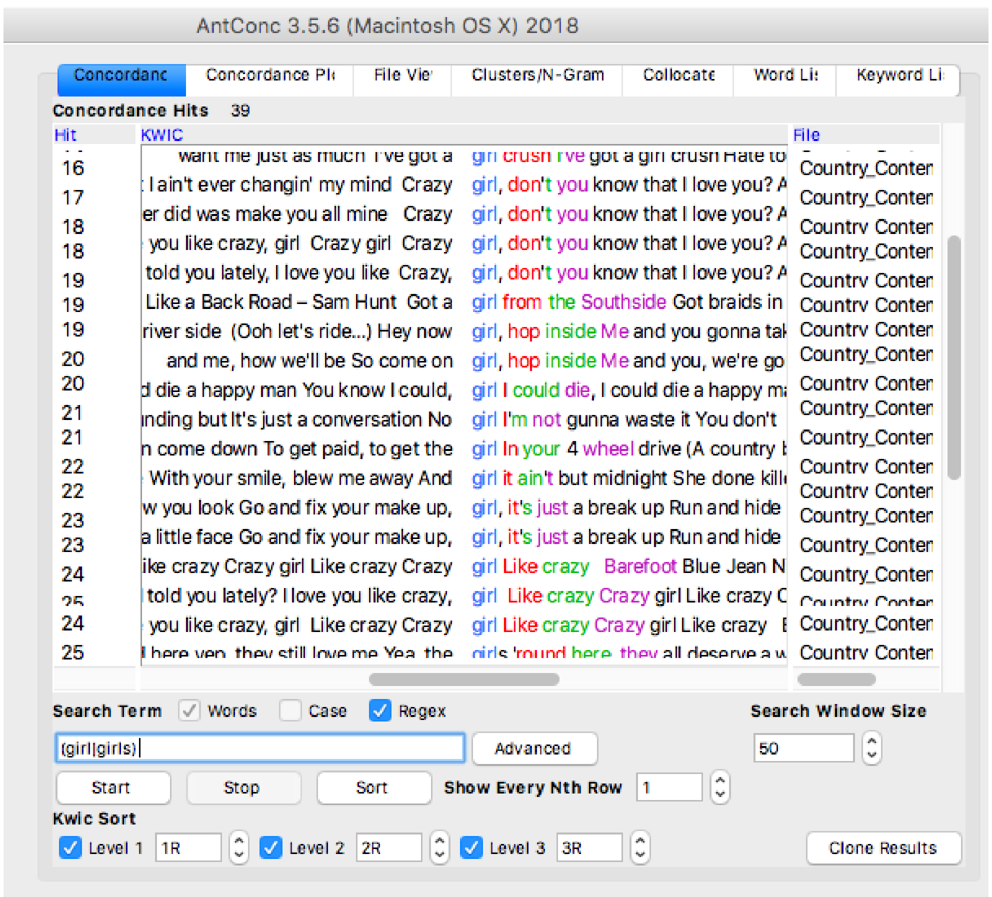


Figure 1. Regular expression search of the gender item *girl*.

Qualitative analysis

We manually categorized lemmas based on Shin (2016) and adapted them towards our data by adding or removing categories as needed. In terms of their reference, the lemmas were categorized as: Partner/Lover, Attractive/Desirable, Male/Female, Adult Male/Female etc. As for their portrayal, the authors categorized the data as follows: Beautiful, Strong/Weak, Romantically Worshiped, Sexually Objectified, Submissive, Insecure. Data were coded by both authors. In Tables 2 and 3 below, we provide examples of the function for each of the categories. These examples serve to provide an overview for how coding functioned in this study.

Table 2

Female portrayal examples

Portrayal	Lyrics
Beautiful	some pretty faces, been with some beautiful **girls**
Faithful	heart of a faithful **woman**
Romantically worshipped/ desired	Always treat your **woman** like a lady
Sexually objectified	**Girl,** look at that body
Submissive	When I walk on by, **girls** be looking like "damn he fly"
Insecure	I was caught somewhere between a **woman** and a child
Strong	I learned something from my blue eyed **girl,** Sink or swim you gotta give it a whirl

Table 3

Male portrayal examples

Portrayal	Lyrics
Confident	all the other **boys** try to chase me But here's my number
Weak	hand in my hand Baby I could die a happy **man**
Desirable	**Boy,** you're my lucky star I wanna walk
Strong	Wanna be your victim, ready for abduction **boy**
Dependent	I can't live without you, **baby**

Undesirable	**Boy** you're an alien, your touch so foreign
Insert	First time I seen her walk by, **Man** I 'Bout fell

Occasionally, lemmas fell into two categories and were coded as such upon agreement of the two authors. For example, in the following example, the word *girl* was coded to portray females as both Beautiful and Romantically Worshiped:

"*...when you smile, the whole world stops and stares for a while 'cause,* ***girl,*** *you're amazing just the way you are*." (Contemporary Pop music)

The categories Beautiful and Romantically worshiped may appear to overlap, but in our study they are considered different concepts. The category Beautiful includes description of physical appearance, while Romantically worshiped illustrates that the speaker adores, desires, or respects the woman, for example: *Always treat your woman like a lady.*

RESULTS

First, we display the frequency for each lemma throughout time and between genres in Table 4 below.

Table 4

Frequency of lemmas throughout time and genres

No. of occurrences	Woman	Girl	Boy	Man	Baby	Total No. of gender items per period
Pop Music 90s		11	20	6	39	76
Contemporary Pop Music	1	37	12		35	84
Country Music 90s	2	16	18	5	15	56
Contemporary Country Music	5	44	8	14	72	143
Total No. of word lemmas	8	108	58	25	161	360

Baby is the most common throughout 90's pop music, while the actual words *man* and *woman* are infrequent. The situation is similar in Contemporary pop music; the use of the lemma *girl* increases in Contemporary pop music and the word *man* completely disappears.

When it comes to Country music, the frequency of the lemmas *girl* and *man* also increased over time. The most surprising finding was the increase across time of the word lemma *baby* from 15, in the 90s, to 72 in Contemporary country music. That suggests that this terminology is used to refer to both *men* and *women* more than the other lemmas in this study.

Gender items in pop and country music - What do they refer to?

In order to respond to RQ1 of what the five lemmas primarily refer to in pop and country music and how have they changed across time and genres, the authors assigned each lemma a category based on how the lemma was used in the context of the lyrics. Then, we assigned each lemma was assigned a category based on its context within the lyric before coding distribution for both music genres.

Table 5

Girl, woman & baby and their reference in ***Pop Music***

	Partner/ Lover		Adult Females		Attractive/ Desirable		Young Female		Friends	
Time Period	90s	Cont.	90s	Cont.	90s	Cont.	90s	Cont	90s	Cont
Girl	9	5	1	1	1	27		1		2
Woman				1						
Baby	5	6		1	14	15				

Table 5 shows the reference for the lemmas *girl, woman,* and *baby* across both decades for Pop music using Shin's (2016) classification. Attractive/Desirable (n=57) was the most frequent category for all gender lemmas in 90s Pop music and Contemporary 90s music across both time periods, and Partner/Lover was second (n=25). These two categories depicted the majority of gender lemma occurrences throughout this genre.

Table 6

Girl, woman & baby and their reference in ***Country Music***

	Partner/ Lover		Adult Females		Attractive/ Desirable		Young Female		Friends	
Time Period	90s	Cont	90s	Cont	90s	Cont	90s	Cont	90s	Cont
Girl	12	22	1	7	2	3	1			6
Woman	1	2				4	1			
Baby	1	52				20				

Table 6 shows what the lemmas *girl*, *woman*, and *baby* referred to across both decades for Country music. Most interestingly, *baby* only occurred once in 90s music while it occurred 72 times in contemporary country music. *Baby* referred to Partner/Lover 52 times and to Attractive/Desirable

20 times. Overall, the lemmas *girl*, *woman*, and *baby* most frequently refer to Partner/Lover (n=100) and Attractive/Desirable (n=29).

Table 7

T*he lemmas boy, man & baby and their reference in **Pop music***

	Adult Males (In General)		Partner / Lover		Authority Figure		Attractive/Desirable Male	
Time Period	90s	Cont.	90s	Cont.	90s	Cont.	90s	Cont.
Boy	4	4	1			3	14	5
Man	1							
Baby			11	4			9	9

Table 7 shows what the lemmas *boy*, *man*, and *baby* referred to across time periods or Pop Music. The lemmas most frequently referred to Attractive/Desirable Male (n=37) and Partner/Lover (n=16). The lemma *baby* most often referred to Attractive/Desirable (n=18) and Partner/Lover (n=15) while the lemma *boy* most often referred to Attractive/Desirable (n=19).

Table 8

*The lemmas boy, man and baby and their reference in **Country music***

	Adult Males		Tough/ Brave		Partner/ Lover		Attractive/Desirable Male	
Time Period	90s	Cont	90s	Cont	90s	Cont	90s	Cont
Boy		10		2	17	1		4
Man	2	1		3		8		
Baby					14			

Table 8 shows what the lemmas *boy, man*, and *baby* referred to across both time periods for Country music. For Country, all lemmas most frequently referred to Partner/Lover (n=40) across decades. The lemma *boy* referred to Partner/Lover 17 times in the 90s and only once in Contemporary music; however, it referred to Adult Male 10 times in Contemporary music and only once in the 90s. Where Pop music (Table 6) most often described men as Partner/Lover or Attractive/Desirable, Country music varied between Partner/Lover and Adult Male.

Female and male portrayal

This section responds to RQ2, how gender is portrayed in pop and country music, and how these portrayals have changed across time. While our first question looked at how gender lemmas referred to men and women, this section looks at how gender lemmas portrayed men and women. Figure 1 illustrates female portrayal for Pop music.

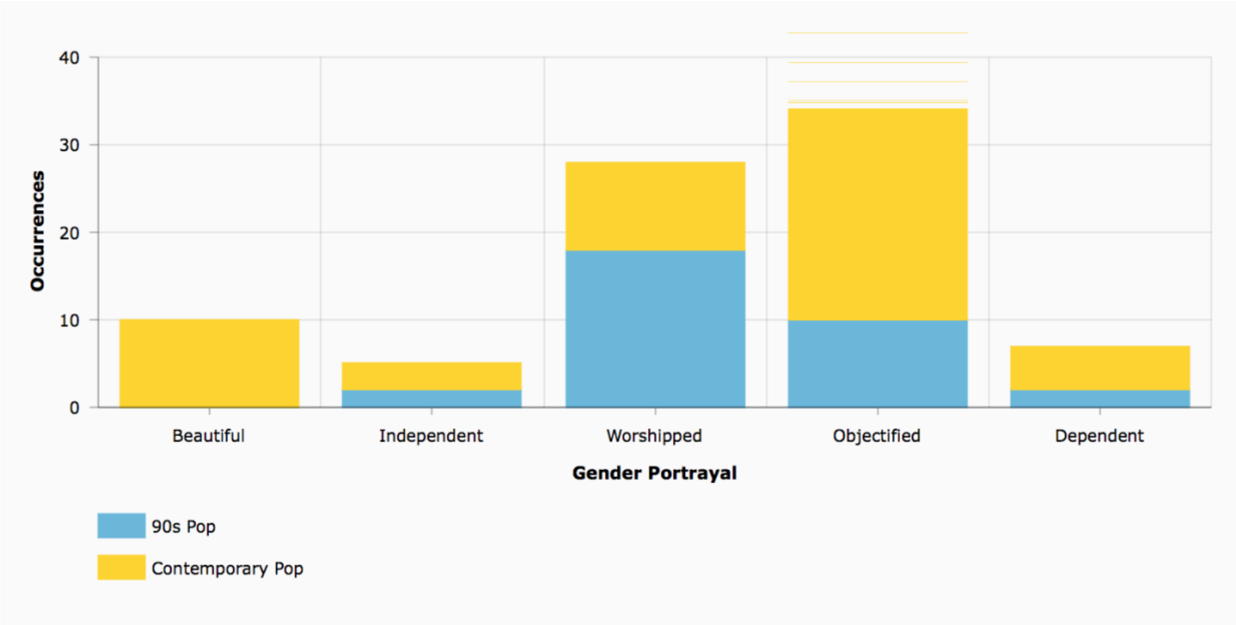


Figure 1. Female portrayal across years for pop music.

In the 90s, females were most frequently portrayed as Romantically Worshiped (n=18) followed by Sexually Objectified (n=10). Whereas in Contemporary pop music, Sexually Objectified occurred 24 times while Romantically Worshipped only occurred 10 times. Contemporary Pop music had more instances of sexual objectification while 90s Pop music had more instances of romantic worship. Figure 2 looks at female portrayal for Country music.

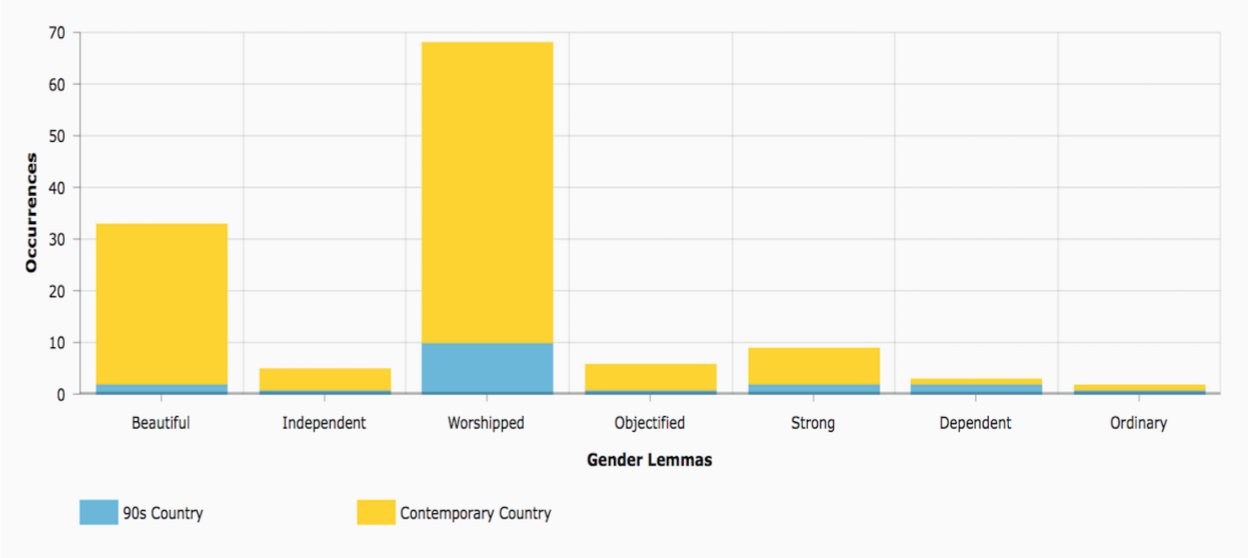


Figure 2. Female portrayal across years for country music.

Romantically Worshipped (n=68) and Beautiful (n=33) were the most frequent referent for women in country music across both time periods. There are considerably more instances of gender lemmas used in contemporary country music, and there does not appear to be clear trends in increase or decrease between time periods.

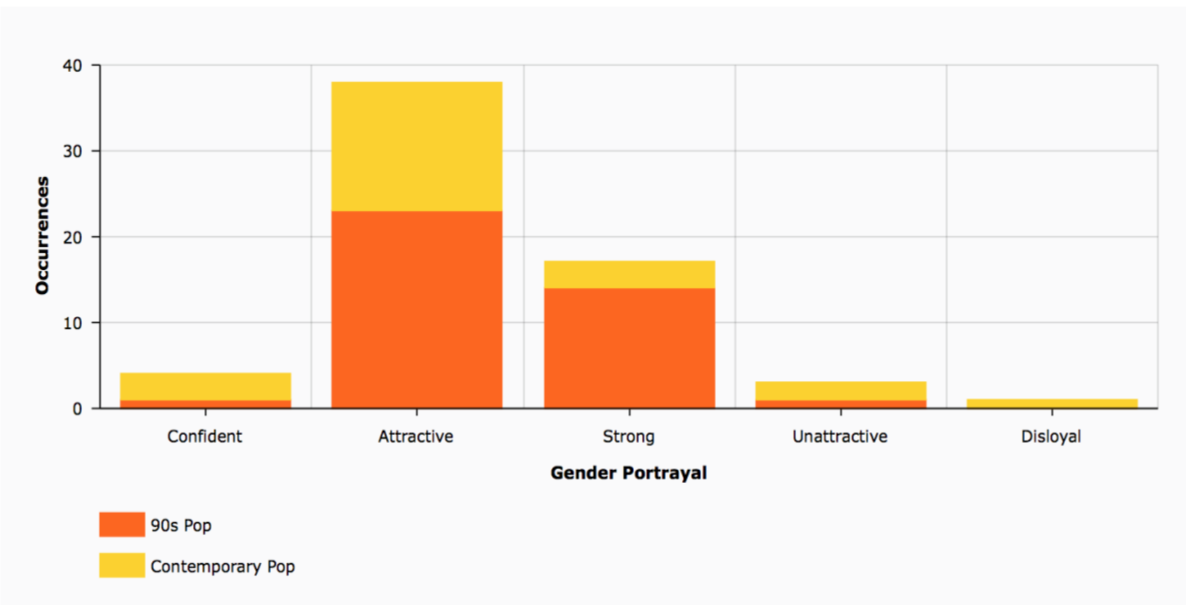


Figure 3. Male portrayal across years for pop music.

Figure 3 illustrates male portrayal for Pop music. During the 90s, men were most often portrayed as Attractive (n=23) and Strong (n=14), and within Contemporary Pop music, men were most frequently portrayed as Attractive (n=15).

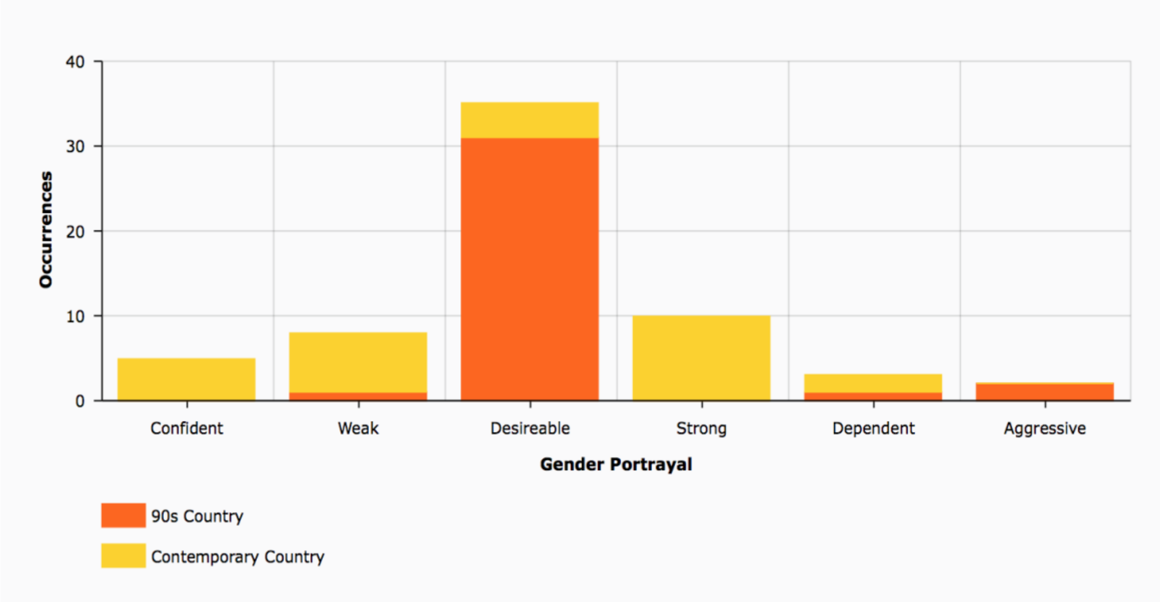


Figure 4. Male portrayal across years for country music.

Finally, Figure 4 depicts male portrayal within Country music. In 90's country, men were most often represented as Desirable (n=31) whereas there was more of a distribution between Confident (n=5), Weak (n=7), Desirable (n=4), Strong (n=10), and Dependent (n=2) in Contemporary Country music.

DISCUSSION

This study investigated how the use of selected gender lemmas changed across time in pop and country music and the way men and women are portrayed within or between music genres. In this section, we discuss the male and female portrayals in these song lyrics and what the results of this study imply.

Female portrayal

Our findings might suggest that representation of female have changed across time and across music genres. Pop music in the 90s mostly used to portray women in the traditional sense as part of families and partners, referring to them as Partner/Lover by using the word lemmas *girl* and *baby*, while Contemporary Pop music mostly portrays them as unofficial partners or object of desire, avoiding the use of the word *woman* in Pop music.

On the other hand, in Country music, we note the opposite. Most of the females are referred to as partners or lovers and their portrayal as Romantically Worshiped indicates the traditional sense of the word lemmas. Nonetheless, things have slightly started to change within Country music as well; in the 90s there were no instances of females being referred to as Attractive/Desirable Female and in Contemporary music the use of this category has started to grow.

Societal changes among the relationships between men and women has started to be reflected in the song lyrics too. Shin (2016) also found that most of the uses of *girl* and *woman* refer to adult females and girlfriends in both Hip-hop and Country music, and unofficial partners only in Hip-hop. While there are certain instances in our data of females being portrayed as Beautiful, Independent, Strong, Submissive, Country music in general mostly portrays females more respectfully compared to Pop music, and as Shin (2016) indicates, in Hip-hop music.

Male Portrayal

The word lemmas used for males in 90s Pop music are mostly referring to Attractive/Desirable people, Partner/Lovers, and Adult Males in general. In Contemporary Pop music, males are mostly referred to as Attractive/Desirable people while other portrayals almost equally present Adult Males, Partners, and Authority Figures. Similar to the female portrayals, there are changes over time where men in Contemporary music are more frequently described as part of nonofficial relationships. Also, males in Pop music are generally presented as attractive, strong, or confident. In addition, the word lemma *man* is rare while *boy* and *baby* are mostly used to refer to men throughout the decades. Males also appear to be portrayed as disloyal by female artists in Contemporary Pop music, which is another indicator that men are not portrayed as part of family in the traditional sense in Pop music.

On the other hand, men in Country music are frequently represented as Partner/Lover throughout both periods. There are instances of men being referred to as Tough/Brave which indicates a more traditional representation of man. In general, males are often portrayed as Attractive/Desirable, but there are some more recent portrayals as both Strong/Weak or Dependent. This finding suggests that males nowadays are presented in various aspects, whereas in the 90s they were mostly viewed as attractive, desirable figures. Contemporary Country had no instances where *baby* referred to males. However, this was may be due to the overwhelming majority of male artists in the top 20 Contemporary Country music songs.

CONCLUSION

Through a linguistic analysis of song lyrics across time and genre, we observed how lyrics language has changed within the American culture. The results suggest that the use of language has changed over time and genres and the scope of these nouns can be much wider than their traditional definitions.

In three instances, we encountered the use of *man* and *boy* as inserts, not referring to anything, for example: *Oh boy, I am so tired.* We excluded those examples from the data above because they were not relevant for our research questions, but this may also be a point worth mentioning to ESL learners. This is an example of a language use that cannot be typically found in the regular textbooks and EFL learners may find this useful.

When it comes to qualitative coding, there were occasionally certain instances that were difficult to interpret and that may belong to two categories at the same time. Personal bias may have influenced the decisions. Future research should consider increasing the corpus size and adding more gender lemmas, such as *lady, madam*, or *mister*, to give a more complete understanding of the genre.

ACKNOWLEDGMENTS

We would like to thank Dr. Gulbahar Beckett for her guidance and support throughout this research.

ABOUT THE AUTHORS

Agata Guskaroska is a Fulbright scholar and MA student in TESL/Applied Linguistics at Iowa State University. She taught ESL courses for 7 years and American literature at FON University. Her major interests include CALL, pronunciation, and SLA.

email: agatag@iastate.edu

Joshua Taylor is a graduate teaching assistant and a second-year PhD student in Rhetoric and Professional Communications at Iowa State University. He has taught a variety of English and Communication classes over the last three years. His major research interest is the socio-political implications of digital rhetorics and language.

Email: jstaylor@iastate.edu

REFERENCES

Anthony, L. (2017). AntConc (version 3.5.6) [computer software]. Retrieved from: http://www.laurenceanthony.net

Baker, P. (2013). Will Ms ever be as frequent as Mr? A corpus-based comparison of gendered terms across four diachronic corpora of British English. *Gender and Language, 4*(1), 125-149.

Barwick, L., Birch, B., & Evans, N. (2007). Iwaidja Jurtbirrk songs: bringing language and music together. *Australian Aboriginal Studies, (2),* 6-34.

Billboard. (2018). Charts archive. Retrieved from https://www.billboard.com

Burton, S. A. (2013). 23 country songs from the '90s that you still know by heart. Retrieved from https://www.buzzfeed.com/summeranne/23-country-songs-from-the-90s-that-you-still-know-by-heart

Daniels, M. (2014). The largest vocabulary in Hip Hop. Retrieved from http://rappers.mdaniels.com.s3-website-us-east-1.amazonaws.com/

Friginal, E. (2007). Outsourced call centers and English in the Philippines. *World Englishes, 26*(3), 331-345.

Hopkins. C. (2011). The hip-hop word count counts on language to understand the reality we keep. Retrieved from https://readwrite.com2011/01/20/the_hip-hop_word_count_counts_on_language_to_under

Motschenbacher, H. (2016). *Language, normativity and Europeanisation: Discursive evidence from the Eurovision song contest*. Basingtoke: Palgrave Macmillan.

Murphey, T. (1990). *Song and music in language learning*. Bern: Peter Lang.

Murphey, T. (1992). The discourse of Pop songs. *TESOL Quarterly, 26*(4), 770-774.

Pearce, M. (2008). Investigating the collocational behaviour of MAN and WOMAN in the British National Corpus using Sketch Engine. *Corpora, 3*(1), 1–29.

Shin, J. (2016). Corpus-based analysis on gendered items in hip-hop and country song lyrics. *Corpus Linguistics Research, 2*, 41.

APPENDIX A.

Top 20 Contemporary/Pop Songs	Top 20 90's Pop Songs
1. Party Rock Anthem – LMFAO 2. We Found Love – Rihanna 3. Somebody That I Used to Know – Gotye 4. Rolling in the Deep – Adele 5. Blurred Lines – Robin Thicke 6. Call Me Maybe – Carly Rae Jepsen 7. Happy – Pharrel Williams 8. Royals – Lorde 9. Dark Horse – Katy Perry 10. Moves Like Jagger – Maroon 5 11. Just The Way You Are – Bruno Mars 12. Thrift Shop – Macklemore & Ryan Lewis 13. One More Night – Maroon 5 14. All Of Me – John Legend 15. We Are Young – Fun 16. Counting Stars – One Republic 17. Radioactive – Imagine Dragons 18. Sexy and I Know It – LMFAO 19. Someone Like You – Adele 20. E.T. – Katy Perry	1. How Do I Live – LeAnn Rimes 2. Macarena – Los Del Rio 3. Un-Break My Heart – Toni Braxton 4. Foolish Games/You Were Meant For Me – Jewel 5. (Everything I Do) I Do It For You – Bryan Adams 6. I'll Make Love To You – Boyz II Men 7. Too Close – Next 8. One Sweet Day – Mariah Carey 9. Truly Madly Deeply – Savage Garden 10. Candle In The Wind – Elton John 11. End of the Road – Boyz II Men 12. The Sign – Ace of Base 13. The Boy is Mine – Brandy and Monica 14. Because I Love You (The Postman Song) – Stevie B 15. Whoomp! (There It Is) – Tag Team 16. Rush Rush – Paula Abdul 17. You're still the one – Shania Twain 18. I Will Always Love You – Whitney Houston 19. Gangsta's Paradise – Coolio 20. Nothing Compares 2 U – Sinead O'Connor

APPENDIX B.

Top 20 Contemporary Country Songs	**Top 20 Country Songs 90's**
1. Love Like Crazy – Lee Brice 2. Why Don't We Just Dance – Josh Turner 3. Crazy Girl – Eli Young Band 4. Barefoot Blue Jean Night – Jake Owen 5. Time Is Love – Josh Turner 6. You Don't Know Her Like I Do – Brantley Gilbert 7. Crusie – Florida George Line 8. Wagon Wheel – Darius Rucker 9. Boys 'round Here – Blake Shelton 10. Get Your Shine On – Florida Georgia Line 11. Mama's Broken Heart – Miranda Lambert 12. This Is How We Roll – Florida George Lina 13. Play It Again – Luke Bryan 14. Take Your Time – Sam Hunt 15. Girl Crush – Little Big Town 16. Highway Don't Care – Tim Mcgraw with Taylor Swift 17. Die A Happy Man – Thomas Rhett 18. Body Like a Back Road – Sam Hunt 19. Hurricane – Luke Combs 20. Sure Be Cool If You Did – Blake Shelton	1. Alan Jackson – Gone Country 2. Trisha Yearwood – She's In Love With The Boy 3. John Michael Montgomery – Life's A Dance 4. Tim McGraw – I Like It I Love It 5. Faith Hill – This Kiss 6. Little Texas – God Blessed Texas 7. George Strait – Check Yes Or No 8. Reba McEntire – Is There Life Out There 9. Shania Twain – You're Still The One 10. Brooks & Dunn – My Maria 11. Dixie Chicks – Wide Open Spaces 12. Martina McBridge – Independence Day 13. Clay Walker – Hypnotize The Moon 14. Mary Chapain Carpenter -- Passionate Kisses 15. Rick Trevino – Bobbie Ann Mason 16. Pam Tillis – Cleopatra, Queen of Denial 17. Vinne Gill – When I Called Your Name 18. Deana Carter – Strawberry Wine 19. Clint Black – Killin' Time 20. Suzy Bogguss – Drive South

Haslam, M., & Zetterholm, E. (2019). The role of consonant clusters in English as a lingua franca intelligibility. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Ames, IA, September 2018 (pp. 276-284). Ames, IA: Iowa State University.

PRESENTATION/POSTER

THE ROLE OF CONSONANT CLUSTERS IN ENGLISH AS A LINGUA FRANCA INTELLIGIBILITY

Mara Haslam, Stockholm University
Elisabeth Zetterholm, Stockholm University

This study attempts to add to the limited body of research on what aspects of English pronunciation affect intelligibility for non-native listeners (users of English as a lingua franca). It addresses the claim from Jenkins' (2000; 2002) Lingua Franca Core that intelligibility will suffer if consonants are deleted from initial consonant clusters or from final clusters in ways that do not fit English phonology, and that addition of extra sounds to a cluster by epenthesis will not adversely affect intelligibility. Monosyllabic words with initial or final consonant clusters produced by talkers of different language backgrounds were played for 11 Swedish listeners, who transcribed what they heard in standard English orthography. Responses were then matched against results of an acoustic analysis. Listeners were more successful overall at identifying the intended word structure than acoustic results would indicate, and this pattern holds for stimuli with final clusters but not those with initial clusters. Deletion of one of the consonants from the cluster was shown to be the most common reason for mismatch. These results partially support the Lingua Franca Core but also demonstrate that the location of epenthesis of a vowel in relation to a consonant cluster affects the likelihood of match or mismatch.

INTRODUCTION

The concept of intelligibility, the ability of "listeners to understand the speaker's intended message" (Derwing & Munro, 2015, p. 1) is important to both researchers and pronunciation teachers. For example, pronunciation teachers would like to know where to focus their efforts: which aspects of pronunciation are most likely to make a difference in understanding? However, intelligibility depends not only on the speaker's pronunciation but also on what the listener finds understandable. Therefore, in order to investigate intelligibility researchers have to think about listeners.

Quite a bit of information is known about what aids intelligibility in English when native speakers are listening. In particular, suprasegmentals have been found to be very important for native-speaker listeners of English (Hahn, 2004; Kang & Pickering, 2011). In the case of English, though, the majority of English users are actually non-native speakers and listeners (Lewis, Simons, & Fennig, 2015). It is therefore likely that many English speakers will find themselves communicating more with non-native speakers than with native speakers, or perhaps not with native speakers at all. It is reasonable to question whether the pronunciation aspects that facilitate intelligibility for native speakers will be the same aspects that facilitate intelligibility for non-native speakers. Pronunciation teachers of English for those who will be using English as a lingua franca are left with few research-based resources upon which to draw.

Jenkins’ (2000, 2002) Lingua Franca Core (LFC) is just the sort of resource that many pronunciation teachers are looking for; it is a syllabus attempting to establish exactly which aspects of pronunciation are important for non-native listeners and which are not important or perhaps might even be detrimental to ELF intelligibility. Many of the points of the LFC are different from what is believed about what facilitates intelligibility for native listeners. For example, the LFC does not include word stress for intelligibility in ELF situations. The LFC is comprehensive, covering both segmental and suprasegmental aspects of pronunciation. As the seminal resource for information on ELF intelligibility, the LFC has been relatively widely published in books for teachers and pronunciation textbooks.

However, the LFC suffers from some methodological issues. In addition to the fact that the data upon which the LFC is based is rather limited, the study was based on listening for situations in ELF conversation that triggered repair and then trying to determine which aspect of the pronunciation had caused the misunderstanding. This method has the potential to miss misunderstandings which were not repaired, such as misunderstandings that the participants did not recognize as misunderstandings or misunderstandings that they chose not to repair. For reasons like these, further research into the LFC has been called for (Dauer, 2003; Haslam & Zetterholm, 2016). In particular, perceptual research has the power to more directly establish the effects of certain pronunciation aspects on intelligibility.

Haslam and Zetterholm (2016) is a first attempt to evaluate one of the aspects of the LFC using perceptual methodology by testing the LFC’s claim that aspiration is required on fortis consonants in initial position in stressed syllables. In the 2016 study perceptual results were compared with acoustic analysis to more directly establish the relationship between acoustic factors and intelligibility for ELF listeners. Results of the study showed that the actual picture of ELF intelligibility is more complicated than the LFC predicted for this situation. Listeners quite successfully identified the target words, regardless of acoustic characteristics. In more detailed acoustic analysis, while the fortis consonants /t/ and /k/ partially followed the LFC’s predictions, results for /p/ showed a completely different pattern. Therefore, these results demonstrate the need for further investigation into the aspects of the LFC for deeper understanding of the picture of ELF intelligibility.

The present study is another investigation into the points of the LFC, specifically the claims about consonant clusters. The LFC requires the following for ELF intelligibility:

“no omission of sounds in word-initial clusters, eg. in promise, string;
“Omission in middle and final clusters only permissible according to L1 English rules of syllable structure, e.g. factsheet can be pronounced ’facsheet’ but not ’fatsheet’ or ’facteet’;…
“Addition is acceptable, for example, ‘product’ pronounced [pər’ɑdʌkʊtɔ] was intelligible to NNS interlocutors, whereas omission was not, for example, ’product’ pronounced /’pɑdʌk/.” (Jenkins, 2002, p. 97)

In the present study, perceptual methodology in combination with acoustic analysis was used to address the following questions:

(1) What strategies do ELF speakers use to pronounce consonant clusters (i.e. deletion, epenthesis, etc.) in initial and final positions?

(2) Do these strategies facilitate or hinder intelligibility for ELF listeners?

METHOD

76 tokens representing 66 unique monosyllabic words with consonant clusters in either initial position (47 words) or final position (36 words) were selected. Recordings of these words were selected from the Wildcat corpus (Bradlow et al., 2007; Van Engen, et al., 2010), specifically from a task where non-native speakers of English who did not share a native language discussed differences between two similar pictures. There were 21 different talkers with the following native languages: Chinese (n=5), Korean (n=8), Persian (n=1), Italian (n=1), Japanese (n=1), Marathi/Hindi (n=1), Russian (n=1), Spanish (n=2), and Thai (n=1).

Listeners were native speakers of Swedish and therefore did not share an L1 with any of the talkers. There were 11 listeners who reported a range of English proficiency from Basic to Advanced.

Listeners were asked to complete a computerized dictation task. For each item, the listener heard a stimulus recording. He/she was then asked to type in what word he/she thought he/she had heard using normal English orthography. Responses were coded for CV structure according to the normal spelling rules of English (e.g. “black” -> CCVC). CV structure of the target word was also coded.

Acoustic analysis was also completed on the stimuli for CV structure using Praat software (Boersma, 2001). Based on the acoustic analysis, each stimulus was assigned a CV structure. CV structures of the responses were then compared to the CV structures from the acoustic analysis and the CV structures of the target words to identify match or mismatch. If the consonant cluster existed in both structures in the targeted position (i.e. initial position or final position), this was counted as match. Therefore, if the target word were “black” (CCVC) and the response were “brag” (CCVC) this would be counted as a match because the CV structure is the same. Mean proportions of match between listeners’ responses and the acoustic analysis were calculated as well as mean proportions of match between listeners’ responses and the CV structure of the target word. When a mismatch occurred, the item was also coded for type of mismatch.

RESULTS

For all stimuli, the mean proportion of match between listeners’ responses and the acoustic analysis was 0.669 (SD=0.03321), while the mean proportion of match between listeners’ responses and the target word was 0.7193 (SD=0.08387). That is, listeners were more successful at identifying the CV structure of the target word than the acoustic analysis would suggest. These results are presented in Figure 1. Results were submitted to paired-samples t-test and the difference between the two means was found to be significant (t(9)=-2.281, p=0.049).

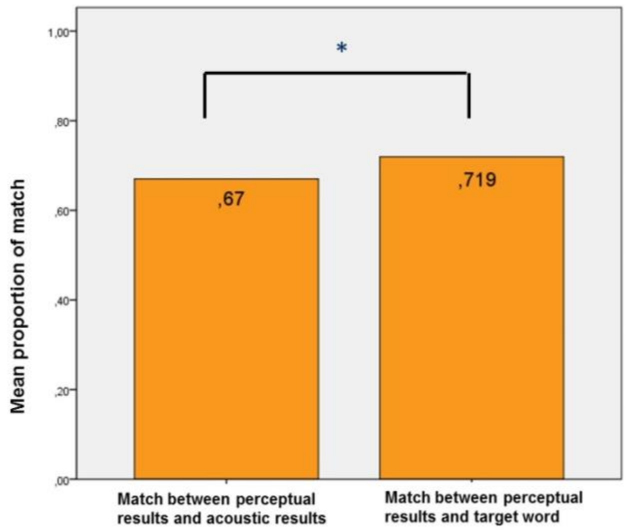


Figure 1. Mean proportion of match between perceptual results and acoustic results, and mean proportion of match between perceptual results and target word for all stimuli.

Additional analysis was conducted for items with clusters in initial position and final position. For clusters in initial position, the mean proportion of match between listeners' responses and the acoustic analysis was 0.7489 (SD=0.04683) while the mean proportion of match between listeners' responses and the target word was 0.7787 (SD=0.10996). These results are visible in Figure 2. While the difference between means indicates that listeners were slightly more successful at identifying the target word than acoustic analysis would suggest, paired t-test results did not find a significant difference in this case (t(9)=-1.121, p=0.291).

For items with clusters in final position, the mean proportion of match between listeners' responses and the acoustic analysis was 0.5667 (SD=0.05885) while the mean proportion of match between listeners' responses and the target word was 0.6417 (SD=0.08013). These results are visible in Figure 3. Paired t-test confirmed that this difference was significant (t(9)=-3.948, p=0.003). Therefore, we can conclude that listeners were more successful at identifying the CV structure of the target word than the acoustic analysis would suggest, even though the overall proportion of match between both listeners' responses and the acoustic analysis and listeners' responses and the target word were lower for final clusters than for initial clusters.

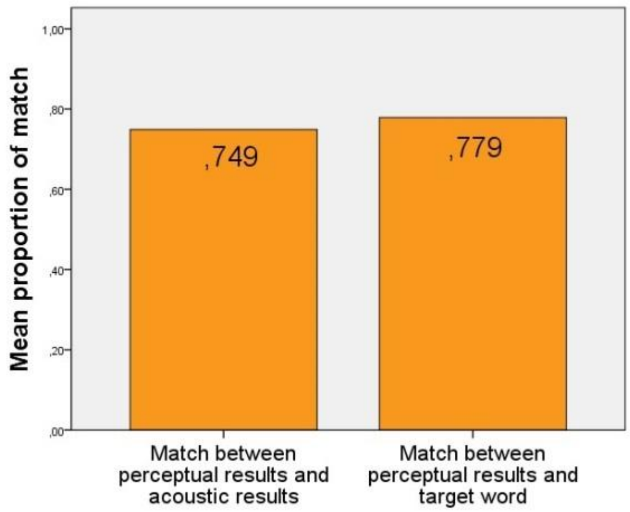


Figure 2. Mean proportion of match between perceptual results and acoustic results, and mean proportion of match between perceptual results and target word for initial clusters.

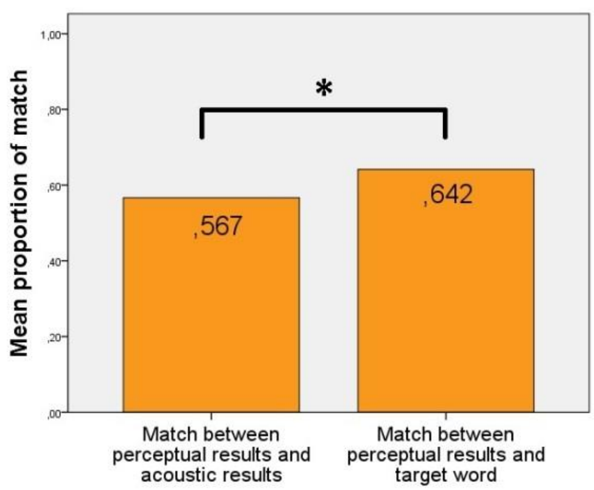


Figure 3. Mean proportion of match between perceptual results and acoustic results, and mean proportion of match between perceptual results and target word for final clusters.

Table 1

Itemized list of reasons for mismatch between acoustic information and perceptual information, separated into acoustic reasons for mismatch and perceptual reasons for mismatch

Acoustic reasons for mismatch		Perceptual reasons for mismatch	
Reason	Number of instances	Reason	Number of instances
Deletion of consonant from cluster	87	Deletion of consonant from cluster	80
Extra initial vowel	31	Extra initial vowel	38
Extra final consonant	9	No response	22
More than 2 extra initial sounds	1	More than 2 extra final sounds	16
		More than 2 extra initial sounds	8
		Extra final consonant	3
		Extra initial consonant	3
		Other differences	2
		Extra final vowel	2
		Extra vowel within consonant cluster	1

Table 2

Itemized list of reasons for mismatch between target word and perceptual information

Perceptual reasons for mismatch	
Reason	Number of instances
Deletion of a consonant from the cluster	127
Extra initial vowel	69
No response	24
2 or more extra final sounds	17
2 or more extra initial sounds	15
Extra final consonant	5
Other differences	3
Extra final vowel	3
Extra initial consonant	2
Extra vowel within consonant cluster	1

A number of different reasons for mismatch between listeners' responses and the acoustic information were identified. For some responses the acoustic analysis revealed deleted or added sounds, while in some cases the reason for the mismatch seemed to lie within the listeners' responses. Reasons for mismatch between the listeners' responses and the target word were also identified. Itemized results of this coding are presented in Tables 1 and 2. In all cases, the lack of a consonant cluster, i.e. deletion of one of the consonants from the consonant cluster, was the top reason for mismatch, followed by addition of an initial vowel to initial consonant clusters.

DISCUSSION

To summarize the results, overall, listeners seemed to perform better at identifying the target word than the acoustic analysis would suggest. This pattern holds for final clusters but not for initial clusters. These results seem to be in line with previous results (Haslam & Zetterholm, 2016) which demonstrated that ELF listeners were often successful at identifying target words regardless of VOT of the initial consonant. In combination with previous results on VOT, the present results may indicate that the particular characteristics of the consonant/s (e.g. VOT) are not especially important for recognition of a word as long as an initial consonant/s is present; that is, listeners may be using top-down skills to recognize words rather than bottom-up processing. This supports Field's (2004) analysis that L2 listeners rely more on top-down processing than bottom-up.

Another area of interest in these results has to do with the relative status of initial clusters vs. final clusters. Listeners performed more accurately in general on recognizing the structure of initial clusters than final clusters. These results can possibly be explained by the fact that some sort of phonetic reduction often happens at the ends of words, such as word-final devoicing produced by some non-native speakers of English (Edge, 1991). However, there was significantly more match between the result of the perceptual test and the target word than there was match between the perceptual result and the acoustic information for final clusters, but there was no corresponding significant difference for initial clusters. These results together indicate that listeners do find final clusters important, but that they may be depending on other information, such as the vowel, to identify words with initial clusters.

The present study was an attempt to support or refute the LFC's points about consonant clusters. These results can be interpreted as partial support of the LFC. The LFC's claim that deletion of consonants from consonant clusters would result in intelligibility seems to be supported: deletion was the top reason for mismatch in all cases. When these results are broken down into initial and final clusters, however, only the final clusters support the LFC's claim.

In addition, the LFC's claim that epenthesized consonant clusters would be intelligible seems to not be supported: the presence of an extra initial vowel (i.e., a vowel inserted before the initial consonant cluster) was the second most common reason for mismatch in these results. Further research can investigate in more detail whether the location of epenthesis is important. In the present study, epenthesis before the initial consonant cluster was found to be a major reason for mismatch, but epenthesis between the consonants of the consonant cluster was a very uncommon reason for mismatch.

Since the present study used only monosyllabic words, further research can also investigate the effect of strategies such as epenthesis and deletion when the consonant cluster is part of a two- or multi-syllable word. As both Haslam and Zetterholm (2016) and the present study suggest, using perceptual methodology to investigate ELF perception can be a valuable line of inquiry. Further research can also focus on other aspects of the LFC such as its claims about suprasegmental pronunciation in addition to segmental aspects.

ABOUT THE AUTHORS

Mara Haslam is a Senior Lecturer in English with language education orientation at the Department of Language Education at Stockholm University. Her research interests include how aspects of pronunciation affect intelligibility in different languages and combinations of speakers and listeners. Her research is inspired by her experiences as a teacher of pronunciation and she hopes that it will help pronunciation teachers to prioritize which aspects of pronunciation they should teach. Mara's teaching efforts focus on helping pre-service teachers prepare to teach English in the Swedish school system. Mara completed a Ph.D. at the University of Utah in 2011 with a dissertation titled *The effect of perceptual training including required lexical access and meaningful linguistic context on second language phonology*.

Elizabeth Zetterholm is a Senior Lecturer and Associate Professor in Swedish as a second language at the Department of Language Education at Stockholm University. Her main research interest is second language pronunciation with a special focus on Swedish. She has been studying how pronunciation is taught in classroom with adult second language learners of Swedish. Another research interest is how orthography can have an impact on pronunciation for second language learners.

REFERENCES

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5*(9/10), 341-345.

Bradlow, A. R., Baker, R. E., Choi, A., Kim, M., & Van Engen, K. J. (2007). The Wildcat corpus of native-and foreign-accented English. *Journal of the Acoustical Society of America*, *121*(5), 3072.

Dauer, R. M. (2005). The Lingua Franca Core: A new model for pronunciation instruction? *TESOL Quarterly*, *39*(3), 543-550.

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

Edge, B. (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. doi:10.1017/S0272263100010032

Field, J. (2004). An insight into listeners' problems: Too much bottom-up or too much top-down? *System*, *32*(3), 363-377.

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

Haslam, M., & Zetterholm, E. (2016). The importance of aspirated initial stops in English as a lingua franca. 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.66-75). Ames, IA: Iowa State University.

Kang, O., & Pickering, L. (2011). The role of objective measures of suprasegmental features in judgments of comprehensibility and oral proficiency in L2 spoken discourse. *Speak Out, 44*, 4–8.

Lewis, M.P., Simons, G.F., & Fennig, C.D. (Eds.). (2015). *Ethnologue: Languages of the world*, (18th edition). Dallas, Texas: SIL International.

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

Van Engen, K. J., Baese-Berk, M., Baker, R. E., Choi, A., Kim, M., & Bradlow, A. R. (2010). The Wildcat Corpus of native-and foreign-accented English: Communicative efficiency across conversational dyads with varying language alignment profiles. *Language and Speech, 53*(4), 510-540.

Hirozane, Y. (2019). Different degrees of effects of pauses on English rate perceived by English and Japanese speakers. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Ames, IA, September 2018 (pp. 285-295). Ames, IA: Iowa State University.

PRESENTATION/POSTER

DIFFERENT DEGREES OF EFFECTS OF PAUSES ON ENGLISH RATE PERCEIVED BY ENGLISH AND JAPANESE SPEAKERS

Yoshito Hirozane, Mejiro University

Pausing is a very important factor when listeners judge the speaking rate. However, pause frequencies are quite different between English and Japanese languages. Roughly speaking, Japanese has three times as many pauses per sentence as English (Anderson-Hsieh & Venkatagiri, 1994). Another difference is that English has a 'rallentando' or a slowing down throughout the intonation phrase (Dankovicová, 1999), while Japanese has mora timing, where every mora is pronounced at approximately the same rate throughout an utterance (Han, 1962; Homma, 1981; Minagawa-Kawai, 1999; Port, Al-Ani, & Maeda, 1980; Sato, 1993). Due to such differences, the degrees of effects of pauses on perceived rate of English could be different. This hypothesis was tested with an experiment. Pairs of English passages, which were identical in physical rate but different in pause frequency, were presented to Japanese speakers and English speakers, who were asked to indicate which passage sounded faster or the same. The results showed that low pause frequency passages tended to be perceived as faster by both Japanese and English speakers. However, there was a higher proportion of Japanese speakers who judged low pause frequency passages as faster in speech rate compared to English speakers. Therefore, pauses appear to have a stronger impact on rate perception by Japanese speakers than English speakers because English speakers take advantage of rallentandos as well as pauses to detect syntactic boundaries while pauses are the only syntactic boundary marker which Japanese speakers can take advantage of.

INTRODUCTION

It is believed that some languages are spoken more quickly than others (Roach, 1998). English is among those languages which are believed to be spoken faster, at least from Japanese listeners' point of view (Griffiths, 1992). Although there might be several reasons that contribute to such beliefs, I propose that the ways pauses function in an utterance play an important role in how Japanese and English speakers perceive English speech rate.

Previous researchers argued that pausing is a very important factor when listeners judge the speaking rate (Den Os, 1988; Feldstein & Bond, 1981; Grosjean & Lane, 1974; Lass, 1970). However, how important the function of pauses as a syntactic boundary maker seem to be different between English and Japanese.

First, the number of pauses in Japanese and English sentences differ significantly. The mean number of pauses per sentence in English spoken by native speakers is zero for short and medium sentences and 0.67 for long sentences (Anderson-Hsieh & Venkatagiri, 1994). On the other hand, the mean number of pauses per sentence in Japanese is 1.8 (Kaiki & Sagisaka, 1996). Roughly speaking, Japanese has three times as many pauses as English.

In addition, pauses are usually placed on major syntactic boundaries (Hawkins, 1971; Ishizaki, 2005; Viola & Madureira, 2008). They tend to function as an indicator of major syntactic boundaries and help listeners understand utterances more easily. Since pauses occur three times as frequently in a Japanese utterance as an English one, it is possible that these pauses serve a more important role for Japanese speakers than English speakers.

English speakers, on the other hand, do not use as many pauses as an indication of syntactic boundaries due to the fact that English has a ‘rallentando’ or a slowing down throughout the intonation phrase (Dankovicová, 1999) (See Figure 1), which also indicates syntactic boundaries.

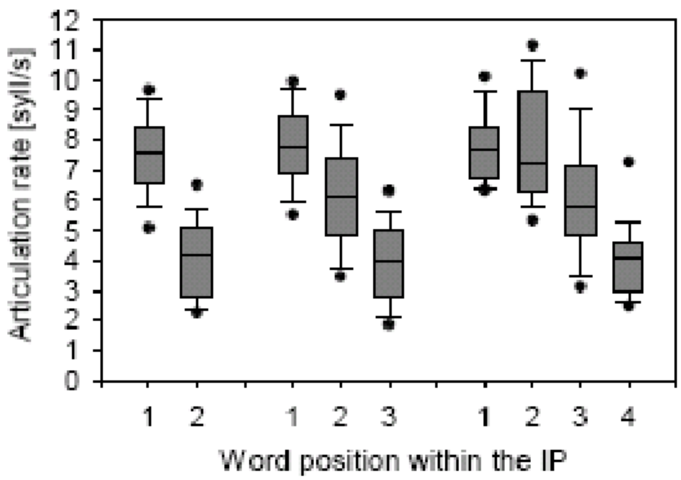


Figure 1. Word articulation rate in different positions within intonation phrase (English) (Dankovicová, 1999).

Considering that short and medium sentences are usually spoken without any pauses (Anderson-Hsieh & Venkatagiri, 1994), the function of pauses as a syntactic boundary marker is not very important in English. Instead, the function is mostly undertaken by rallentandos.

Research question

Since the function of pauses as a syntactic boundary marker is not the same in its importance between English and Japanese, it can be hypothesized that pauses affect perception of rate differently between English speakers and Japanese speakers. My research question is: Do pauses have a stronger impact on rate perception by Japanese speakers than English speakers?

METHODS

Participants

Twenty-two native Japanese speakers (4 males and 18 females) participated in the experiment. They were all undergraduate students of a university in Tokyo, Japan, all of whom majored in English. Their English proficiency was at a lower intermediate level on average. As a control group, 25 native English speakers (25 females) participated in this study. They were all undergraduate students of a university in Sydney, Australia. None of them had studied Japanese as a foreign language prior to the experiment. None of the participants had any hearing loss or hearing impairment.

Although Feldstein, Dohm, and Crown (1993) concluded that females tend to judge speech rates to be faster than men do and the results of this study are subject to be potentially biased for the English speakers due to the fact that they were all female, I would argue that the nature of the task will eliminate such doubt. In their experiment, the task was to estimate the rates of speech on a 7-point scale. In this experiment, however, the task was to compare rates, not to estimate rates as will be explained in the procedure section. Hence, even if females overestimate the rates of a pair of tokens, they could fairly compare pairs of tokens and judge which is faster or the same just the same way as male participants would do. Comparison of rates by females will not be biased due to their overestimation of rates compared to males.

Stimuli

Fifteen English passages, previously used for the speaking section in the Test in Practical English Proficiency (EIKEN) Grade 3, were selected as the materials for the speech stimuli. Each passage was composed of three sentences. The “EIKEN” is a leading language assessment in Japan and Grade 3 is recognized as a benchmark proficiency level for junior high school graduates, which is equivalent to CEFR level A1. Although all passages in the speaking section of the Grade 3 tests are supposed to be easily understood by Japanese college students in general, comparatively easy passages were selected as the materials for the tokens to exclude parsing effects as much as possible. If they are too difficult to understand, the participants may start wondering whether they are judging rate or comprehensibility. All the stimuli were synthesized with the Festival Speech Synthesis System (Black & Clark, 2003).

Two types of pause treatment were implemented to each passage. For the first type of treatment, passages had two intersentence pauses only, the length of which was around 1000ms. For the second type of treatment, passages had three intrasentence pauses in each sentence in addition to the two intersentence pauses. The duration of the intrasentence pause was fixed at 150ms while that of the intersentence pause was at 300ms. Prior to the pause distribution modification, the speech parts of the paired passages had been controlled so that both had the same duration. Then, the two versions of the same passage with different treatment of pauses were paired together to be presented to the participants. Thus, paired passages had exactly the same speech duration and pause duration. The only difference was pause distribution. One had only two pauses and the other had 10 to 12 pauses.

The locations of the pauses were carefully planned as well. Pauses after a long sequence of continuous speech will add more cognitive load on memory than those after a short speech stream due to the word-length effect on working memory (Baddeley, Thomson, & Buchanan, 1975). In turn, variations of memory load could affect rate perception. Hence, to control for an equal cognitive load on memory for the listeners, the intrasentence pauses were placed so that the intervals may roughly be the same. In addition, although natural speech could have been used with modifications such as pause insertion and deletion, it was not used because such modifications would possibly result in unnaturalness due to discontinuous intonation contours.

An example of the paired passage tokens is shown below with the locations of the pauses marked. Intersentence pauses are indicated by <<P>> and intrasentence pauses are indicated by <p>.

Beach passage

The sea is a popular place to go in summer. <<P>> Some people enjoy sunshine on the beach and others play in the cool water. <<P>> Everyone can have a good time together.

The sea <p> is a popular place <p> to go <p> in summer. <<P>> Some people <p> enjoy sunshine on the beach <p> and others <p> play in the cool water. <<P>> Everyone <p> can have <p> a good time <p> together.

Procedures

For the Japanese participants, the experiment was conducted on one participant at a time in a quiet room using Praat (Boersma & Weenink, 2009). During the experiment, the participant was seated in front of a computer screen showing three rectangles lined up horizontally which were labelled “1st”, “same”, “2nd” from left to right meaning respectively “the first stimulus sounds faster than the second one”,” both sound the same”, and “the second stimulus sounds faster than the first one”. Fifteen pairs of English passages, one of which was a high pause frequency passage and the other was a low pause frequency passage, were randomly presented twice in different orders to the participant over headphones with an interval of 0.5s between the paired stimulus passages. The participant was asked to indicate which sequence of a given pair sounded faster or if both sounded the same in terms of speech rate by clicking one of the three rectangles on the screen. The order of presentation of the stimulus pairs were counterbalanced across the participants. The total number of trials was 30 for each participant.

For the English speakers, the experiment was conducted in small groups of two or three in the perception experiment room at a university in Sydney, Australia. The experiment was administered on paper. The stimulus passages were presented to the English speakers over headphones just the same way as to the Japanese speakers except that the order of presentation of the stimulus pairs was not counterbalanced across the participants. They were asked to indicate their responses on a sheet of paper by circling one of the three options (“1st faster”, “same”, “2nd faster”) printed on it. The three options corresponded to the three response rectangles presented to the Japanese speakers.

Different experimental designs were used for Japanese speakers and English speakers due to the time constraints on my visit to Australia. The experiments were conducted on groups of Australian subjects rather than individually to collect more data in shorter time. In addition, no computer software was used for the Australian subjects because no software was available for groups of subjects.

RESULTS

The binomial test was carried out for an alpha level of 0.05 to determine the effect of frequency of pauses on Japanese and English listeners' perception of English speech rates. The results show that in most cases Japanese speakers perceived the low pause frequency passages as faster than the high pause frequency passages, as shown in Table 1, where 11 out of 15 results tested significant. English speakers also perceived the low pause frequency passages as faster than the high pause frequency passages in some cases, but not as often as Japanese speakers as shown in Table 2, where seven out of 15 results tested significant.

Table 1

Number of Japanese speakers' responses having indicated that the low pause frequency passage was faster, the high pause frequency passage was faster, and both sounded the same

Passage name	Low pause frequency	High pause frequency	same	Test results
beach	19	8	17	* (p= .026)
bike	26	5	13	** (p< .001)
cleaning	19	12	13	
cleanup	24	6	14	** (p< .001)
cooking	23	8	13	** (p= .005)
earth	27	10	7	** (p= .004)
fastfood	14	16	14	
foreign	15	8	21	
internet	26	7	11	** (p< .001)
library	25	5	14	** (p< .001)
picture	22	10	12	* (p= .025)
present	13	9	22	
rainy	27	5	12	** (p< .001)
river	28	7	9	** (p< .001)
spring	24	10	10	* (p= .012)
Total	332	126	202	

Table 2

Number of English speakers' responses having indicated that the low pause frequency passage was faster, the high pause frequency passage was faster, and both sounded the same

Passage name	Low pause frequency	High pause frequency	same	Test results
beach	17	16	17	
bike	15	21	14	
cleaning	19	15	16	
cleanup	28	10	12	** (p= .003)
cooking	13	17	20	
earth	27	11	12	** (p= .007)
fastfood	20	14	16	
foreign	14	19	17	
internet	32	10	8	** (p< .001)
library	26	12	12	* (p= .017)
picture	28	13	9	* (p= .014)
present	17	18	15	
rainy	29	8	13	** (p< .001)
river	36	9	5	** (p< .001)
spring	26	16	8	
Total	347	209	194	

Then, a chi-square test of goodness-of-fit was performed for an alpha level of 0.05 to determine whether the low and high pause frequency passages were equally judged to be faster. The Japanese speakers' responses were not equally distributed, $\chi 2$ (14, N=458) = 111.41, $p < .01$. The English speakers' responses were not equally distributed, either, $\chi 2$ (14, N=556) = 71.81, $p < .01$. The results indicate that the low pause frequency passages were judged to be faster than the high pause frequency passages by both the Japanese and the English speakers.

Furthermore, the 2-sample test for equality of proportions revealed that there was a significant difference between the proportions of the Japanese and English speakers who judged that the low pause frequency passage was faster than the high pause frequency passage ($p < .001$). The proportion of the Japanese speakers who judged that the low pause frequency passage was faster was greater than that of the English speakers.

DISCUSSION

Although contribution of pause frequency to rate perception was already reported by Grosjean & Lane (1974) and Grosjean & Lane (1976), there was a methodological flaw in their studies, as increasing/decreasing the number of pauses meant decreasing/increasing the speaking rate. To address this issue the paired passages used in the present experiment had exactly the same articulation rate and speaking rate in terms of any unit commonly used for rate measurement. Nonetheless, the results show that the low pause frequency passages were perceived as faster than the high pause frequency passages. Also the ratio of Japanese speakers who judged the low pause frequency passages as faster was higher than that of English speakers. In other words, two passages having exactly the same physical rate could be perceived as different in rate by Japanese and English speakers when pauses are distributed differently, and the frequency of pauses has a stronger impact on rate perception by Japanese speakers than English speakers. Why?

Japanese speakers' expectation and memory span

People try to listen to a second language through the ears of a first (Cutler, 2000). Since intrasentence pauses function as a major clue for Japanese speakers to detect syntactic boundaries in Japanese, Japanese speakers may well expect pauses to occur as often as Japanese even when they listen to English.

Therefore, without pauses Japanese speakers may experience difficulties in identifying the syntactic boundaries in English. They would have more difficulties understanding the structure of a sentence and its meaning, which causes processing delay. Moreover, memory span is shorter in a foreign language than in the native language (Lado, 1965). This would further slow down Japanese speakers' English processing rate. Although processing delay may be compensated for by insertion of the intrasentence pauses, they do not appear very often in English. Japanese speakers may try to make up for the processing delay by taking advantage of a few intersentence pauses. Yet, more often than not, before they are able to do so, the next sentence goes into their ears without mercy. That could be one of the reasons why Japanese speakers perceive English as faster when there are fewer pauses.

Rallentando advantage

On the other hand, English speakers may take advantage of the presence of rallentandos as a marker of syntactic boundaries, using intrasentence pauses as a secondary boundary marker. Although the intrasentence pauses also provided information about syntactic boundaries, their contribution to comprehension was not very great for English speakers because they were already given enough information about syntactic boundaries by virtue of rallentandos.

The high pause frequency passages had longer sentences (including pauses) than the low pause frequency passages because the latter had no intrasentence pauses. The high pause frequency passages, on the other hand, had shorter intersentence pauses (300 ms) compared to those of the low pause frequency passages (1000 ms). According to Lass (1970), intersentence pause time alterations show greater changes in rate perceived by English speakers than intrasentence pause time alterations. Frequent intrasentence pauses decreased perceived rate but the shorter intersentence pause time may have increased the perceived rate. Hence, intra- and intersentence pauses may have cancelled each other in their effects on perceived rate by English speakers.

Japanese speakers, on the other hand, may have tried very hard to detect syntactic boundaries especially when they listened to the passages without intrasentence pauses. Unlike English speakers, however, Japanese speakers lack the ability to identify syntactic boundaries through rallentandos. Intrasentence pauses then become the sole cues for Japanese speakers to take on to detect syntactic boundaries.

Why no rallentando in Japanese?

The higher pause frequency could be a language-specific feature of Japanese, which is not just a habit or a favorite style of Japanese way of speaking but a logical consequence of the timing feature of Japanese.

Rallentando is not allowed in Japanese because it is a mora-timed language. Oono and Miwa (1996) measured the sentence initial and sentence final mora durations of Japanese read by a professional narrator. They found that the durations were the same at both sentence initial and sentence final positions. Japanese speakers usually pronounce every mora at approximately the same rate throughout an utterance. Slowing down would disturb the mora-timing.

There has been a lot of controversy over whether the Japanese mora is really isochronous or not (Warner & Arai, 2001). While much evidence has been reported against isochrony of the Japanese mora (Beckman, 1982; Campbell & Sagisaka, 1991; Hoequist, 1983; Otake, 1988, 1989), evidence for isochrony has been reported as well by many researchers (Han, 1962; Homma, 1981; Minagawa-Kawai, 1999; Port et al., 1980; Sato, 1993). It would be impossible to achieve complete isochrony in any language. Although Japanese speakers are not speaking like a metronome and durational variation is acceptable to a certain degree, slowing down of a Japanese sentence at the end of each intonational phrase would modify the time component of Japanese to the extent that it does not sound natural. In an extreme case, it might cause misunderstanding because long and short vowels are distinctive in Japanese: ho “canvas”, hoo “cheek”, hoooo “pontiff”. Since Japanese is not allowed to slow down, it chose to insert pauses from time to time instead.

CONCLUSIONS

Since English speakers locate syntactic boundaries mostly using rallentandos, pauses are not always necessary for them to process the sentence. Adding more pauses may not help English speakers understand an utterance better.

On the other hand, Japanese speakers may not be familiar with rallentandos in English and their function as a syntactic boundary marker. Therefore, pauses may help Japanese speakers detect syntactic boundaries more easily while the presence of rallentandos may not help at all. With fewer pauses, Japanese speakers would have more trouble understanding the structure of an utterance, which causes processing delay and makes it sound faster. Thus, my hypothesis was confirmed: pauses have a stronger impact on rate perception by Japanese speakers than English speakers.

ABOUT THE AUTHOR

Yoshito Hirozane (PhD, Macquarie University, 2013) is a Professor of English and linguistics at Mejiro University, Tokyo. His research interest includes L2 speech rate perception, listening comprehension teaching, and pronunciation teaching.
Contact information:
Mejiro University
4-31-1 Naka-ochiai Shinjuku-ku
Tokyo 161-8539 Japan
hirozane@mejiro.ac.jp

ACKNOWLEDGEMENTS

An earlier version of this article was included as part of the author's PhD dissertation "Perception of English rate by Japanese L2 learners of English" submitted to Macquarie University. Preparation of this article was supported by JSPS KAKENHI Grant Number 17K02980.

REFERENCES

Anderson-Hsieh, J., & Venkatagiri, H. (1994). Syllable duration and pausing in the speech of Chinese ESL speakers. *TESOL Quarterly, 28*(4), 807-812.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14*(6), 575-589. doi:10.1016/S0022-5371(75)80045-4

Beckman, M. (1982). Segment duration and the 'Mora' in Japanese. *Phonetica, 39*(2-3), 113-135.

Black, A. W., & Clark, R. (2003). The Festival Speech Synthesis System (Version 1.4.3). Retrieved from http://www.cstr.ed.ac.uk/projects/festival/

Boersma, P., & Weenink, D. (2009). Praat: doing phonetics by computer (Version 5.1.05). Retrieved from http://www.praat.org/

Campbell, N., & Sagisaka, Y. (1991). Moraic and syllable-level effects on speech timing. (Onsei taimingu ni mirareru moora to onsetsu no eikyou ni tsuite) "音声タイミングにみられるモーラと音節の影響について ". Trans. Committee Speech Res., *Acoust. Soc. Japan 『電子情報通信学会技術研究報告』, SP90*, 35-40.

Cutler, A. (2000). Listening to a second language through the ears of a first. *Interpreting, 5*(1), 1-23. doi:10.1075/intp.5.1.02cut

Dankovicová, J. (1999). Articulation rate variation within the intonation phrase in Czech and English. *ICPhS99*, 269-272.

Den Os, E. A. (1988). Rhythm and tempo of Dutch and Italian: A contrastive study. (PhD), University of Utrecht, The Netherlands.

Feldstein, S., & Bond, R. N. (1981). Perception of speech rate as a function of vocal intensity and frequency. *Language and Speech, 24*(4), 387-394. doi:10.1177/002383098102400408

Feldstein, S., Dohm, F.-A., & Crown, C. L. (1993). Gender as a mediator in the perception of speech rate. *The Bulletin of the Psychonomic Society, 31*(6), 521-524.

Griffiths, R. (1992). Speech Rate and Listening Comprehension: Further Evidence of the Relationship. *TESOL Quarterly, 26*(2), 385-390.

Grosjean, F., & Lane, H. (1974). Effects of two temporal variables on the listener's perception of reading rate. *Journal of Experimental Psychology, 102*(5), 893-896. doi:10.1037/h0036323

Grosjean, F., & Lane, H. (1976). How the listener integrates the components of speaking rate. *J Exp Psychol Hum Percept Perform, 2*(4), 538-543.

Han, M. S. (1962). The feature of duration in Japanese. *Onsei no Kenkyuu, 10*, 65–80.

Hawkins, P. R. (1971). The syntactic location of hesitation pauses. *Language and Speech, 14*(3), 277-288. doi:10.1177/002383097101400308

Hoequist, C. E. (1983). Parameters of speech rate perception. In K. J. Kohler & C. E. Hoequist (Eds.), *Studies in speech timing* (Vol. 20, pp. 99-138). Kiel: Inst. für Phonetik.

Homma, Y. (1981). Durational relationship between Japanese stops and vowels. *Journal of Phonetics, 9*, 273–281.

Ishizaki, A. (2005). Nihongo no ondoku ni oite gakushusha wa donoyoni pozu o okuka 日本語の音読において学習者はどのようにポーズを置くか. Sekai no nihongo kyoiku 世界の日本語教育, 15, 75-89.

Kaiki, N., & Sagisaka, Y. (1996). Study of pause insertion rules based on local phrase dependency structure. *IEICE Trans., J79-D-II*(9), 1455-1463.

Lado, R. (1965). Memory span as a factor in second language learning. *IRAL - International Review of Applied Linguistics in Language Teaching, 3*(2), 123-130. doi:10.1515/iral.1965.3.2.123

Lass, N. J. (1970). The significance of intra- and intersentence pause times in perceptual judgments of oral reading rate. *J Speech Hear Res, 13*(4), 777-784.

Minagawa-Kawai, Y. (1999). Preciseness of temporal compensation in Japanese mora timing. Paper presented at the *Proceedings of the 14th International Congress of Phonetic Science*.

Oono, M., & Miwa, J. (1996). Rodoku ni okeru pozu to hatsuwa sokudo 朗読におけるポーズと発話速度. *The Journal of the Center for Educational Practice, Research and Training at Iwate University Education Department『岩手大学教育学部附属教育実践研究指導センター研究紀要』*(6), 45-58.

Otake, T. (1988). A temporal compensation effect in Arabic and Japanese. *Bull. phonet. Soc. Japan, 189*(19-24).

Otake, T. (1989). A cross linguistic contrast in the temporal compensation effect. *Bull. phonet. Soc. Japan, 191*, 14-19.

Port, R. F., Al-Ani, S., & Maeda, S. (1980). Temporal compensation and universal phonetics. *Phonetica, 37*(4), 235-252.

Roach, P. (1998). Some languages are spoken more quickly than others. In L. Bauer & P. Trudgill (Eds.), *Language Myths* (pp. 150-158). London: Penguin.

Sato, Y. (1993). The durations of syllable-final nasals and the mora hypothesis in Japanese. *Phonetica*, *50*(1), 44-67.

Viola, I. C., & Madureira, S. (2008). The roles of pause in speech expression. *Fourth International Conference on Speech Prosody* (pp. 721-724). Campinas, Brazil.

Warner, N., & Arai, T. (2001). Japanese mora-timing: A review. *Phonetica, 58*(1-2), 1-25.

Jarosz, A. (2019). Secondary school learners' pronunciation needs, perceptions and attitudes. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Ames, IA, September 2018 (pp. 296-306). Ames, IA: Iowa State University.

PRESENTATION/POSTER

SECONDARY SCHOOL LEARNERS' PRONUNCIATION NEEDS, PERCEPTIONS AND ATTITUDES

Anna Jarosz, University of Łódź, Poland

A longitudinal action-research study was conducted in a secondary school context. It aimed to explore learners' pronunciation needs and to assess the impact of 30 weeks of pronunciation instruction on learners' perceived phonetic competence, awareness and self-confidence. The results suggest that this group of learners can display a considerable degree of linguistic awareness, which they felt was developed through the course of instruction. Although the instructed pronunciation training focused mainly on segmental features, it contributed to the learners' convictions that they improved their pronunciation and became more conscious speakers.

INTRODUCTION

Pronunciation constitutes a significant aspect of oral communication. One of the most relevant issues in the field of pronunciation instruction is the dichotomy between nativeness and comfortable intelligibility as learning goals stressed by Levis (2005). With the dominance of the Communicative Approach, the nativeness principle seems to have been abandoned for the sake of intelligibility. Derwing and Munro (2015) explain the terms intelligibility (comprehension) and comprehensibility (effort put into understanding) and their lack of correlation with accentedness (the degree of L1 accent in L2 speech). Another dilemma in the current research on pronunciation instruction is the relationship between accuracy, fluency and proficiency in the language. Accuracy is most frequently identified with segmental correctness and the degree of approximation to the model phonemic category (Waniek-Klimczak, 2018). Fluency of speech, on the other hand, refers to widely-understood suprasegmental phonetic features such as rhythm, stress or intonation, whereas the linguistic proficiency in its broadest meaning refers to the organisation of speech on all levels (i.e., both segmental and suprasegmental; Waniek-Klimczak, 2003). Derwing and Munro (2015) characterise fluency as fluidity which is "the degree to which speech flows easily without pauses and other dysfluency markers" (p. 5). Even though the terms proficiency and fluency are sometimes used interchangeably, Derwing and Munro (2015) identify proficiency as the highest level of fluency distinguished by creative use of the language (Fillmore, 1979).

To date, much research in the field has concentrated on teaching and learning English as second language (ESL) (Derwing, Diepenbroek, & Foote, 2012; Derwing & Munro, 2011; Derwing, Munro, & Thomson, 2007). With regard to English as foreign language (EFL), many studies conducted in Poland focus on university-level students of English and their attitudes towards pronunciation instruction (Pawlak, Mystkowska-Wiertelak, & Bielak, 2015; Sobkowiak, 2002; Waniek-Klimczak, 2013; Waniek-Klimczak, Rojczyk, & Porzuczek, 2015). Huensch and Thompson (2017) report that a relatively small number of studies concentrate on the attitudes of learners toward pronunciation in an FL context (their study focuses on adult American-university FL learners' pronunciation attitudes). In Europe, Smit and Dalton (2000) and Smit (2002) investigated motivations and self-efficacy of adult German EFL learners. Sardegna, Lee, and Kusey (2014, 2018) undertook to explore the motivations and beliefs regarding English pronunciation of adolescent Korean learners. Using the LAMP Inventory (Learner Attitudes and Motivations for Pronunciation five-point Likert-scale survey), the authors found that the learners exhibited a low degree of cognitive, conative and self-efficacy attitudes towards learning English phonology (2014), and that higher self-efficacy affected positively pronunciation skills and strategy use (2018). As was observed by Sardegna, Lee, and Kusey (2014), the research dedicated to motivation and perceptions of young adult learners, however, is scarce in the European EFL contexts and the field needs further exploration and investigation given that there is a growing demand for global communication which relies largely on oral skills, such as accuracy and fluency of speech. Another important reason for selecting younger learners is that they are generally considered more successful at FL pronunciation learning than adults (Singleton & Ryan, 2004). Therefore, it seems reasonable to claim that a population in need of additional study are secondary school students learning English as a foreign language. This group is limited in their exposure to the foreign language (they learn English in Poland and the use of external widely-available materials depends on their choice and will) and they are usually not far enough along in their educational pursuits that they have decided on a career choice. This study, therefore, aims to bridge the existing gap in the literature by describing a longitudinal action research project which occurred with secondary school students aged 17 and 18 who received a year of weekly pronunciation instruction.

The current study

The pronunciation instruction comprised 30 lessons (45 minutes each) of phonetic training as an extracurricular course for 10 self-selected participants. The course design reflected pronunciation syllabi guidelines outlined in literature concerning the functional load principle (Catford, 1987), setting realistic and attainable goals (Morley, 1994; Scheuer, 2015), or the cyclical nature of curriculum development (Derwing & Munro, 2015). Furthermore, studies conducted in the Polish context indicate specific areas of difficulty for Polish learners of English, such as aspiration, distinction between long/short vowels, velarized /l/, dental fricatives, velar nasal, unstressed syllable/vowel reduction, rhythm, stress timing and linking (Porzuczek, Rojczyk, & Arabski, 2013; Rojczyk & Porzuczek, 2012; Szpyra-Kozłowska et al., 2002; Szpyra-Kozłowska, 2005; Szpyra-Kozłowska, 2015; Wells, 2005). Dental fricatives are usually of interest to Polish learners since they do not occur in the Polish sound inventory. Even though they do not affect intelligibility (low functional load), they are salient features of English phonetics and their mispronunciations are frequently perceived as irritating (Scheuer, 2003). In view of the above-mentioned research, the aspects selected for the pronunciation training comprised mainly segmentals (long/short vowel contrast, schwa /ə/ and trap vowel /æ/, aspiration, pre-fortis clipping, final devoicing, dental fricatives /θ ð/, velar nasal /ŋ/) as well as selected suprasegmentals such as rhythm, stress, weak forms and linking. The dominance of segmental phonetics in the pronunciation course resulted from the needs indicated by the participants themselves in their pre-course questionnaires. The range of techniques and training methods involved both the traditional 'listen and repeat' tasks, minimal pairs practice, word-level accuracy targeting word stress and single sounds, homophones (which were often pronounced differently) reading out loud and acting out dialogues, as well as tongue twisters, pronunciation jokes and games like 'bingo' or 'hangman' and the use of pronunciation smartphone applications (Celce-Murcia et al., 2010; Kelly, 2000). Explicit pronunciation instruction was provided with the use of metalanguage, gradually introduced during the course.

Research questions

The study explores secondary school learners' needs and attitudes with regard to features of pronunciation and observes the development of pronunciation awareness and speaking confidence resulting from an extended period of pronunciation instruction. The motivation for the study stemmed from long-term observations of English teaching in the Polish state institution context (which point to inadequacy or lack of pronunciation instruction) as well as conviction that pronunciation constitutes a vital factor affecting speech (Pawlak et al., 2015; Waniek-Klimczak & Klimczak, 2005) and a linguistic skill that merits SLA attention (Gilbert, 2010; Grant, 2014). Furthermore, it was surmised that secondary school learners realise that pronunciation is an essential element of language learning (Baker, 1992; Szyszka, 2015; Tergujeff, 2013). In addition, contrary to coursebooks' authors, it seemed reasonable to expect advanced, relatively fluent and conscious speakers of English to wish to improve their accuracy so as to sound more native-like rather than phonologically simplified (e.g., Jenkins' LFC, 2000). Consequently, the study was guided by the following research questions:

> RQ1: What do secondary school EFL learners perceive their pronunciation needs to be?
> RQ2: Does regular pronunciation instruction lead to attitude changes regarding pronunciation awareness, speaking confidence, and perceived performance enhancement?

METHODS

Participants

The participant group included ten learners aged 17-18, two males and eight females in their second year, who chose an extended-level English course (seven hours of English weekly), plus the additional hour of phonetic instruction. They were all fluent from B1 (intermediate) to B2+ (advanced intermediate) according to the Common European Framework of Reference. The participants had clearly-set learning goals and future plans. Seven of them wanted to pursue a career in a medicine-related field, three in journalism, but none planned to study English. All exhibited a considerably high degree of linguistic awareness since they volunteered for the course, voiced what they felt their pronunciation needs and deficiencies were and wished to achieve satisfaction with their speech while reducing speaking anxiety. They emphasised the desire to acquire more confidence in out-of-class interactions with both native and non-native speakers.

Instruments

This study employed three questionnaires (open-ended pre-, open-ended mid- and Likert-scale post-course), interviews, observations and pre- and post-course speech assessment of *Please call Stella* (speecharchive.gmu.edu). The article will focus on detailed results of the pre- and post-course questionnaires.

The pre-course questionnaire asked the participants to share their opinions and beliefs on five areas: their English background, how they understood the term 'correct' pronunciation, their preferred accent model, the importance of pronunciation in foreign language learning and their pronunciation goals and needs. The end-course questionnaire contained 25 statements on a 6-point Likert scale, where 6 meant definitely agree and 1 meant definitely disagree, with space for additional comments. It aimed to elicit the learners' reflections and their evaluation of perceived development as well as efficacy and practicality of the course. The questionnaires differed in form and scope since they were intended to serve a different purpose: in the first one the participants expressed general opinions and needs whereas in the other (more detailed and more metalinguistic in its nature) they evaluated the course and the knowledge acquired. Both were anonymous and both were administered in Polish. The data analysis procedure for the pre-course questionnaire consists in reporting all response patterns (with the number in brackets representing the frequency of comments), though not all 10 participants responded to all sections. The results of the post-course questionnaire are presented with the mean value and standard deviation for each statement. Additional comments made by some participants are presented as well.

RESULTS AND ANALYSIS

Pre-course questionnaire

All participants started learning English in kindergarten and later they all attended different extra-curricular English courses, 4 in the primary school and 6 in the junior secondary school. However, none of them attended extra English instruction outside the upper secondary context. In their attempts to define correct pronunciation, they mentioned: correct articulation of sounds (n=1), correct word stress (n=1), care in pronouncing words (n=1), accuracy (n=3), native-like speech (n=4) and fluency (n=4). As for their preferred accent variant, three pointed to British English and one to American English. As regards the relevance of pronunciation, they highlighted the communicative aspect of speech and the dominant role of pronunciation (n=2) as well as pronunciation as the most important subsystem of language (n=2). Some of them also asserted that pronunciation errors may lead to misunderstandings and misinterpretations (n=3). The questionnaire shed light on the participants' perceived needs, as follows:

- intelligibility in English-speaking countries (n=1)
- native-like pronunciation and not being recognized as a non-native speaker (n=1)
- acquiring correct pronunciation (n=1)
- reducing their Polish accent while speaking English (n=3)
- fluency in speech (n=4)
- word and sentence stress (n=4)
- accuracy (n=4)

Both fluency and accuracy were indicated as main goals by four (different) learners. When mentioning accuracy, the participants stressed the need to produce sounds accurately, especially English sounds that do not exist in the Polish sound inventory. Only one person evoked intelligibility ('being understood') as the communication objective, whereas four pointed to accent-free or native-like pronunciation ('speaking like natives') as their aim.

Post-course questionnaire

Table 1 presents the questionnaire and results by providing mean value and standard deviation for each statement/question.
Table 1.

The mean and SD values for the end-course questionnaire

Statements	Mean value	Standard deviation
1. I consider the additional English phonetics classes useful.	5.6	0.66
2. I observe improvements to my pronunciation.	4.9	0.53
3. My pronunciation awareness is raised now.	5.6	0.48
4. I pay attention to correct pronunciation while I speak.	5.4	0.48
5. Correct pronunciation is crucial when we speak a foreign language.	5.8	0.40
6. I like to repeat out loud words with difficult sounds.	4.9	1.30
7. I regard phonetic transcription as necessary.	4.7	1.48
8. Knowing phonetic transcription helps read new words in a good way.	5.1	0.94
9. The aspects of connected speech presented during the course were new for me.	4.2	0.97
10. The awareness of connected speech processes is useful when we learn English.	5.0	0.63
11. Word stress is essential for correct pronunciation.	4.8	0.60
12. Sentence stress is essential for correct pronunciation.	4.8	0.74
13. Pronunciation of 'th' was difficult for me.	4.0	1.34
14. Pronunciation of 'th' is now easy.	4.5	0.67
15. I had problems with the velar nasal sound.	3.5	1.36
16. I can pronounce the velar nasal now.	4.8	0.74
17. English vowels are difficult to pronounce.	3.5	1.20
18. I am more aware of the English vowels now.	5.0	0.63
19. Pronunciation of 'ash' was difficult for me (9 answers only).	3.7	1.44
20. Pronunciation of 'ash' is now easier (9 answers only).	4.8	1.08

21. I was not aware that voicing at the end of words is indicated by the preceding vowel length.	4.7	1.26
22. I have learnt that schwa is the shortest vowel in English.	5.0	0.89
23. Did the additional phonetics course help you learn English?	5.5	0.50
24. Would you like to continue the pronunciation course next year?	5.4	0.66
25. Is phonetics relevant in foreign language learning?	5.6	0.48

As can be observed in Table 1, the first five statements reflect the participants' perceptions regarding their self-awareness of pronunciation and the role pronunciation fulfils in the learning process and in speech. Relatively low values of SD (in statements 1-5) indicate that most of the participants' answers varied between the 'strongly agree' and 'agree' options. Statements 6, 7 and 8 referred to the strategies employed while learning pronunciation, i.e. repetition and phonetic transcription. The highest value of SD in statement 7 was caused by one 'definitely disagree' answer. The next fourteen statements denote the progress as perceived by the learners themselves. They all claimed to have mastered pronunciation of difficult segments. Interestingly, there is a discrepancy between the learners' self-beliefs about their progress and their actual performance. Despite their assertions, the recordings' data analysis demonstrates that erroneous pronunciation still occurred. The participants also agreed that word stress, sentence stress and connected speech are essential for pronunciation and contribute to enhanced communicativeness. The results of the last three questions suggest that the participants regarded phonetic training as relevant in foreign language learning in general. They unanimously volunteered to participate in the course continuation next year, which may support the conclusion that their raised awareness stimulated them to wish to learn more.

Seven participants provided additional comments listed below:

- *I like the fact that we speak a lot, even though my results are not always the best. I find it much easier now to speak English in general.*
- *It is difficult to say whether my pronunciation has improved but, definitely, I am more aware of phonetics and I pay more attention to how I pronounce words.*
- *I want to learn more 'correct pronunciation' and more transcription.*
- *I want to pronounce words accurately.*
- *I would like to learn reading from transcription.*
- *I need more everyday fast speech practice.*
- *Now I really pay attention to how I pronounce words.*

These reflections lead to a tentative conclusion that perceptually the course made the participants more sensitive to various aspects of pronunciation thus rendering their speech more controlled and less accidental. It seems they knew what they wanted to improve and work on in the future in order to reach their goals, even though in the Likert-scale they claimed progress in segmental features.

DISCUSSION

In regards to the research questions, the analysis of the learners' needs and expectations proved a relatively high degree of linguistic and phonetic awareness. The study was conducted in response to the popular demand of the learners, who had voiced their needs in the field of pronunciation. They realised pronunciation is crucial in language learning, especially in the light of its communicative function. The participants assessed their capacity and pinpointed their weaknesses and the areas that needed further improvement. With clearly-set needs and expectations, they lacked, however, the requisite metalanguage to explicitly elaborate on them. Apart from fluency, they mentioned accurate articulation of English sounds as their goal, which demonstrates dissatisfaction with their oral production and aspirations for overall correctness and not just rough approximation to the target model. The results of the post-course questionnaire clearly suggest the participants' growing and developing awareness of pronunciation, pronunciation strategies' application and metalinguistic competence.

Furthermore, in their comments, the learners reported becoming more sensitive to speech and its conscious monitoring. They realised they had been equipped with necessary tools leading to the enhancement of their speaking and communication efficacy and the metalinguistic knowledge required to express their needs. While the outcome might not be surprising, it undoubtedly strengthens the argument that pronunciation must be taught and learners wish to learn it.

It can be inferred that secondary school learners have the potential to develop increased phonetic awareness and this potential should not be neglected. Thus, learner needs should be analysed and accounted for in language syllabi. Moreover, contrary to common beliefs of coursebooks' authors, language proficiency increases sensitivity to the accuracy of speech. Therefore, segmental pronunciation practice should not be relegated from the upper- and higher levels; it could be offered parallelly with suprasegmental features. Polish students, similarly to others in Europe (Tergujeff, 2013), realise that too little time is devoted to phonetic instruction in school. Pronunciation should be taught at all educational stages to avoid fossilisation of errors (Baker, 1992).

Regular and planned phonetic training (albeit not very intensive and devoted mainly to segments) exerts a positive influence on the participants' perceptions and self-beliefs regarding their own speech. The data (statement 5) comply with the hypothesis that pronunciation is a relevant factor affecting speech and communication efficacy (Pawlak et al., 2015; Waniek-Klimczak & Klimczak, 2005). Furthermore, contrary to Gimson (1970) and Jenkins (2000) phonological simplifications do not necessarily constitute the learners' ultimate goal. Some of them firmly believe that the beauty of the language lies in its unique pronunciation features.

CONCLUSION

In conclusion, it bears pointing out that the ten participants, despite different interests, varied levels of advancement and school grades, had one common feature, namely their wish to master English pronunciation accurately, which resulted in high attendance and eagerness to dedicate their free time and as a result contributed to the success of the study. Given the limited scope of research concerning state school context, the project has certain important pedagogical implications in that students wish to improve their pronunciation and perceive it as a crucial factor in communication.

ABOUT THE AUTHOR

Anna Jarosz, PhD, is a research assistant at the Department of English Language and Applied Linguistics at the University of Łódź, Poland. She is an experienced teacher of English and specialises in phonetics and phonology as well as teacher training. Her main areas of interest are: FL pronunciation acquisition, pronunciation pedagogy and phonetics. Apart from the English Studies, she also received an MA diploma in Italian.

Contact information:
Department of English Language and Applied Linguistics, University of Łódź
Pomorska 171/173, 91-404 Łódź,
Poland
Email: anna.jarosz@uni.lodz.pl

REFERENCES

Baker, A. (1992). *Introducing English pronunciation: A teacher's guide to tree or three? And ship or sheep?* Cambridge: Cambridge UP.

Catford, J. C. (1987). Phonetics and the teaching of pronunciation. In J. Morley (Ed.), *Current perspectives on pronunciation: Practices anchored in theory.* Alexandria, VA: Teachers of English to Speakers of Other Languages.

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

Derwing, T. M., Munro, M. J., & Thomson, R.I. (2007). A longitudinal study of ESL learners' fluency and comprehensibility development. *Applied Linguistics*, *29*, 359-380.

Derwing, T. M., & Munro, M. J. (2011). Intelligibility, comprehensibility and accent: Their relevance to pronunciation teaching. *Speak Out! IATEFL Pronunciation Special Interest Group Newsletter,* (X), 1-5.

Derwing, T. M., Diepenbroek, L. G., & Foote, J. A. (2012). How well do general-skills ESL textbooks address pronunciation? *TESL Canada Journal*, *30*(1), 22-44.

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

Fillmore, C. J. (1979). On fluency. In C. J. Fillmore, D. Kempler, & W. S. -Y. Wang (Eds.), *Individual differences in language ability and language behaviour* (pp. 173-183). New York: Academic Press.

Gilbert. J. (2010). Pronunciation as orphan: What can be done? *Speak Out! IATEFL Pronunciation Special Interest Newsletter*, *43*, 3-7.

Gimson, A. C. (1970). *An introduction to the pronunciation of English* (2nd ed.). London: Edward Arnold.

Grant, L. (2014). Prologue to the myths: What teachers need to know. In L. Grant with D. M. Brinton, T. Derwing, M. J. Munro, J. Field, J. Gilbert, J. Murphy, R. Thomson, B. Zielinski & L. Yates (Eds.), *Pronunciation myths: Applying second language research to classroom teaching* (pp. 1-33). Ann Arbor, MI: University of Michigan Press.

Huensch, A., & Thompson, A. S. (2017). Contextualizing attitudes toward pronunciation: Foreign language learners in the US. *Foreign Language Annals*, *50*(2), 410-432. doi:10.1111/flan.12259

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

Jenkins, J. (2003). *World Englishes*. London: Routledge.

Kelly, G. (2000). *How to teach pronunciation*. Malaysia: Pearson Longman.

Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, *39*, 369-377.

Morley, J. (1994). *Pronunciation pedagogy and theory, new views, new directions*. Alexandria: TESOL.

Pawlak, M. (2006). On the use of pronunciation learning strategies by Polish foreign language learners. In W. Sobkowiak & E. Waniek-Klimczak (Eds.), *Dydaktyka fonetyki języka obcego w Polsce. Referaty z szóstej konferencji naukowej* (pp. 121- 135). Mikorzyn, 8-10 maja 2006. Konin: Państwowa Wyższa Szkoła Zawodowa w Koninie.

Pawlak, M. (2008). Another look at the use of pronunciation learning strategies. An advanced learner's perspective. In E. Waniek-Klimczak (Ed.), *Issues in accents of English* (pp. 304-322). Newcastle upon Tyne: Cambridge Scholars Publishing.

Pawlak, M., Mystkowska-Wiertelak, A., & Bielak, J. (2015). Exploring advanced learners' beliefs about pronunciation instruction and their relationship with attainment. In E. Waniek-Klimczak & M. Pawlak (Eds.), *Teaching and researching the pronunciation of English. Studies in Honour of Włodzimierz Sobkowiak* (pp. 3-22). Switzerland: Springer.

Rojczyk, A. & Porzuczek, A. (2012). Selected aspects in the acquisition of English phonology by Polish learners – Segments and Prosody. In D. Gabryś-Barker (Ed.), *Readings in second language acquisition* (pp. 93-120). Katowice: Wydawnictwo Uniwersytetu Śląskiego.

Porzuczek, A., Rojczyk, A., & Arabski, J. (2013). *Praktyczny kurs wymowy angielskiej dla Polaków*. Katowice: Wydawnictwo Uniwesytetu Śląskiego.

Sardegna, V. G., Lee, J., & Kusey, C. (2014). Development and validation of the learner attitudes and motivations for pronunciation (LAMP) inventory. *System*, *47*, 162-175. doi: 10.1016/j.system.2014.10.009

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. doi:10.1111/lang.12263

Scarcella, R. & Oxford, R. L. (1994). Second language pronunciation: State of the art in instruction. *System*. *22*(2), 221-230.

Scheuer, S. (2003). What to teach and what not to teach? Some reflections on the relative salience of interlanguage errors. In: Sobkowiak, W. & Waniek-Klimczak, E. (eds). *Dydaktyka fonetyki języka obcego. Zeszyty Naukowe Państwowej Wyższej Szkoły Zawodowej w Koninie 2* (pp. 93-99). Konin: Państwowa Wyższa Szkoła Zawodowa w Koninie.

Scheuer, S. (2015). What to teach and what not to teach, yet again: On the elusive priorities for L2 English phonetics. In E. Waniek-Klimczak & M. Pawlak (Eds.), *Teaching and researching the pronunciation of English. Studies in honour of Włodzimierz Sobkowiak* (pp. 139-152). Switzerland: Springer.

Singleton, D. M., & Ryan, L. (2004). *Language acquisition: The age factor*. Clevedon, UK: Multilingual Matters.

Smit, U., & Dalton, C. (2000). Motivation in advanced EFL pronunciation learners. *International Review of Applied Linguistics in Language Teaching*, *38*, 229-246.

Smit, U. (2002). The interaction of motivation and achievement in advanced EFL pronunciation learners. *IRAL*, *40*, 89-116.

Sobkowiak, W. (1996). *English phonetics for Poles*. Poznań: Bene Nati.

Sobkowiak, W. (2002). English speech in Polish eyes: What university students think about English pronunciation teaching and learning. In E. Waniek-Klimczak & J. P. Melia (Eds.), *Accents and speech in teaching English phonetics and phonology: EFL perspective* (pp. 177-196). Frankfurt: Peter Lang.

Szpyra-Kozłowska, J., Frankiewicz, J., & Gonet, W. (2002). Aspekty fonetyki angielskiej nauczane w polskich szkołach średnich. In W. Sobkowiak & E. Waniek-Klimczak (Eds.), *Dydaktyka fonetyki języka obcego. Neofilologia. Tom II* (pp. 9-27). Płock: Zeszyty Naukowe PWSZ w Płocku.

Szpyra-Kozłowska, J. (2005). Lingua franca core, phonetic universals and the Polish context. In K. Dziubalska-Kołaczyk & J. Przedlacka (Eds.), *English pronunciation models: A changing scene* (pp. 151-176). Bern: Peter Lang.

Szyszka, M. (2015). Good English pronunciation users and their pronunciation learning strategies. *Research in Language, 13*(1), 93-106.

Tergujeff, E. (2013). Learner perspective on English pronunciation teaching in an EFL context. *Research in Language, 11*(1), 81-96.

Waniek-Klimczak, E., & Klimczak, K. (2005). Target in speech development: Learners' views. In K. Dziubalska-Kołaczyk & J. Przedlacka (Eds.), *English pronunciation models: A changing scene* (pp. 229-249). Bern: Peter Lang.

Waniek-Klimczak, E. (2013). Teaching English pronunciation to English majors in Poland: Selected observations. In E. Mańczak-Wohlfeld & M. Jodłowiec (Eds.), *Exploring the microcosm and macrocosm of language teaching and learning* (pp. 229-242). Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.

Waniek-Klimczak, E., & Klimczak, K. (2005). Target in speech development: Learners' views. In K. Dziubalska-Kołaczyk & J. Przedlacka (Eds.), *English pronunciation models: A changing scene* (pp. 229-249). Bern: Peter Lang.

Waniek-Klimczak, E., Rojczyk, A., & Porzuczek, A. (2015). 'Polglish' in Polish eyes: What English studies majors think about their pronunciation in English. In E. Waniek-Klimczak & M. Pawlak (Eds.), *Teaching and researching the pronunciation of English. Studies in honour of Włodzimierz Sobkowiak* (pp. 23-34). Switzerland: Springer.

Wells, J. C. (2005). Goals in teaching English pronunciation. In K. Dziubalska-Kołaczyk & J. Przedlacka (Eds.), *English pronunciation models: A changing scene* (pp. 101-110). Bern: Peter Lang.

Wrembel, M. (2002). Miejsce fonetyki języka angielskiego w szkole – implikacje dla kształcenia nauczycieli. In W. Sobkowiak & E. Waniek-Klimczak (Eds.), *Dydaktyka fonetyki języka obcego. Neofilologia II* (pp. 29-40). Płock: Zeszyty Naukowe PWSZ w Płocku.

Lorenzen, J. (2019). Tonal recall: Musical ability and toneme recognition. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Ames, IA, September 2018 (pp. 307-317). Ames, IA: Iowa State University.

PRESENTATION/POSTER

TONAL RECALL: MUSICAL ABILITY AND TONEME RECOGNITION

Jane Lorenzen, University of North Texas

Because music and language both involve aural patterns and segments, they are speculated to share cognitive processes. Previous studies into the possible link between language-learning ability and musical ability have had inconsistent results. The current project investigated a possible correlation between musical ability and phonological perception. Specifically, this project addressed the following research questions: Is there a correlation between musical memory and the ability to recognize Mandarin tonemes? If so, does this correlation differ by gender or age? Adult participants completed a tone-deafness assessment that measured their pitch perception and short-term musical memory. Participants then watched a YouTube video that explained the tones used in Mandarin. After watching the video, they played an online game to measure their ability to identify the Mandarin tones that they had just learned. Finally, participants recorded their scores and demographic information (language history, age, and gender) in an online survey. The correlation between participants' musical and toneme scores was calculated. The analysis showed a positive correlation between the two scores, with a slightly higher correlation for men than for women, and a higher correlation for older participants. The results suggest that musical ability is one of the individual differences that might confer a slight advantage on some second-language learners.

INTRODUCTION

Many people believe that musicians are better than non-musicians at learning languages. Speculation about links between music and language has occurred for nearly a century (Nakata, 2002), but little is known about what those proposed links entail. This project investigates a potential link between specific aspects of musical ability, namely pitch perception and melodic memory, and one aspect of language-learning ability, phonological perception.

It is logical to expect a link between musical and language-learning abilities. Music and language are both forms of cultural expression, and they both comprise hierarchical structures in which smaller units (notes and phonemes) make up larger units (chords and musical phrases; morphemes and words), which in turn make up even larger units (melodies and sentences).

Both music and language rely on listeners' predictions. According to Gibson's (2000) Dependency Locality Theory, listeners predict certain syntactic categories based on the words they have already heard (for example, in English, after hearing a determiner followed by an adjective, the listener expects a noun or another adjective, but not a verb), and listeners make a connection between an incoming word and the words that have preceded it (Patel, 2003). In a similar way, music listeners who hear certain chords have an expectation, based on the music's key signature and the preceding chord progression, of a limited set of chords to "resolve" the musical phrase. This expectation is evidenced by research showing that "the processing of a musical target is faster and more accurate when it is harmonically related to the preceding prime context" (Tillmann et al., 2003, p. 145).

Some music and language functions have been found to use common parts of the brain. Listening to music activates the same brain regions (the temporal and fronto-temporal areas of both the left and right hemispheres) as language comprehension and production. There is evidence that neuroplasticity is involved in skill acquisition in both musical instrument playing and second language pronunciation, as both skills benefit from acquisition before puberty (Milovanov & Tervaniemi, 2011). In a 2002 study, participants who heard unexpected notes in chord sequences showed activation in the Broca and Wernicke areas, previously thought to be used for language processing only (Koelsch et al., 2002).

LITERATURE REVIEW

Until recently, relatively little of the research into individual differences in language learning addressed the potential relationship between music and language. Skehan (1989), for example, examined diverse individual differences (intelligence, age, anxiety, learning styles, introversion, and risk-taking) but did not address musical ability. More recent studies have had mixed results in showing a link between musical ability and language-learning ability. The following studies illustrate the complexity of this proposed relationship.

Peynircioğlu, Durgunoğlu, and Öney-Küsefoğlu (2002) found a link between musical ability and *phonological awareness*, a broad set of skills in recognizing and manipulating individual sounds in speech, including rhythm, rhyming, alliteration, and sound substitution. They conducted two experiments—one with Turkish-speaking children and one with English-speaking children—to determine whether children who scored high on musical aptitude tests would also score better on tasks that required them to delete phonemes from words and pseudo-words and to delete tones from familiar melodies. None of the children, aged six and younger, could read at even the most elementary level; influence from written forms of the words was thus prevented.

The researchers used data from the highest-scoring and lowest-scoring participants; participant scores closer to the middle of the scale were not included. Predictably, children who scored higher on the musical aptitude test did better at the tone deletion test (omitting notes from familiar songs). These same students also scored higher on the phoneme-deletion tasks (omitting certain sounds from words). The researchers inferred that auditory processing of phonology and music used the same analytical skills and that the similar findings in two very different languages suggest cognitive universals rather than language-specific features.

Nakata (2002) studied the relationship between musical abilities and language discrimination. Nakata did not examine other individual differences that might correlate with musical or language abilities.

Nakata used only adult participants, reasoning that any link between musical ability and language-learning ability in participants who were past the *critical period* (the period, corresponding roughly to puberty, after which phonological acquisition becomes markedly more difficult) would provide a stronger predictor of learners' ability to acquire a new language's phonology. All participants were native English speakers with little exposure to the Japanese language and no formal music training.

Nakata hypothesized that certain specific musical abilities would correlate with specific language skills: recognizing syncopation would correspond to identifying geminates (doubled consonants) in Japanese, recognizing musical time durations would correspond to identifying lengthened Japanese vowels, and recognizing musical pitch would correspond to identifying Japanese pitch accentuation.

Participants were assessed on musical abilities that corresponded to language perception and production tasks. Assessments of recognition of syncopation, suspended notes, and pitch differences corresponded, respectively, to tasks involving Japanese geminates (/kita/ ‘arrived’ vs. /kitta/ ‘sliced’), lengthened vowels (/obasan/ 'aunt' vs. /obaːsan. 'grandmother'), and spoken pitch differences (/kaꜜki/ ‘oyster’ vs. /kakiꜜ/ ‘fence’).

The results of the Nakata study were somewhat surprising. Participants’ mean scores were similar for the paired music and language tasks, but the overall scores on both language and music tests were unexpectedly high. A correlation of .370 was found between musical rhythm recognition and geminate production, but no such correlation was found between musical ability and phonological perception. Nakata theorizes that this split might suggest separate cognitive operations for perception and production. Nakata’s study concedes that study results might have been different if the tests in that study had used both short-term and long-term memory to gauge participants’ language abilities.

Slevc and Miyake (2006) conducted a robust study of the relationship between musical ability and language-learning ability. 50 adult Japanese L1 learners of English took an extensive battery of tests of musical abilities, language-learning abilities, general intelligence, length of residence in the U.S., age of arrival, motivation, and extent of English exposure. All participants resided in the United States and arrived after the age of 11.

Researchers assessed and collected data about participants’ English skills, nonverbal intelligence, phonological short-term memory, musical ability (chord analysis, pitch change identification, tonal memory, and tonal production), and language history (age of arrival, length of U.S. residence, extent of exposure, and motivation). In phonological perception tests, participants heard recorded words and sentences and were asked to distinguish between minimal pairs (for example, “flee’ and “free”). In production tests, participants read aloud a story and a series of minimal-pair words, which were judged by native speakers.

Slevc and Miyake found that musical ability correlated significantly with phonological perception (Pearson correlation coefficient .52) and with phonological production (.45). In the case of phonological perception, musical ability correlated more closely than any other individual difference measured in the study. In the case of production, musical ability correlated similarly to age of arrival, extent of exposure, and length of residence.

Hierarchical regression controlled for the influence of other individual differences in the analysis of the music-language relationship. Slevc and Mikaye confirmed that musical ability predicted differences in phonological perception and production. Their study provided empirical confirmation of the music-language link. From the results of their study, they theorized that any ability that aids in the analysis of sounds is “likely beneficial” in adult second language learning.

The most recent and extensive study, by Bowles, Chang, and Karuzis (2016), controlled for general language-learning ability to examine whether pitch ability can predict tonal word learning. Native English-speaking young adults took cognitive tests, foreign language aptitude tests, and musical aptitude tests. They took pitch ability tests, incorporating both linguistic and non-linguistic tones, to determine their ability to discriminate between tones, identify tones, determine whether two pitches were the same, and identify pitch contours.

Participants learned a set of Mandarin pseudo-words on which they were later retested. This retesting after an interval (up to two days) addressed a deficit in the Nakata study, which tested participants immediately.

Bowles et al. (2016) found stronger correlations than Nakata (e.g., a .440 correlation between one test of pitch ability and a test of phonological short-term memory). Tonal word learning in Mandarin was predicted by participants’ linguistic pitch processing (ability to differentiate between spoken pitch contours) and musicality assessments. The strongest predictor, however, was the measurement of linguistic pitch processing, not musicality. The researchers inferred that nonverbal pitch processing skills contribute to the initial learning of tone by speakers of nontonal languages. They found that musicality, while it may facilitate tone distinction in beginning tonal language speakers, does not seem to confer any advantage on speakers once they have learned the basic tone distinctions.

Based on these studies, we cannot generalize a link between overall musical ability and the broad set of skills associated with language learning. More likely, there are specific musical abilities that predict equally specific aspects of language learning. It is also possible that the specific musical abilities and aspects of language learning vary based on the phonology of the language being learned. The current study looked specifically at pitch recognition and musical memory to determine whether they correlated with toneme recognition in Mandarin.

Research questions

This project investigated a possible correlation between musical ability and phonological perception. Specifically, the study addressed these research questions:

1. Is there a correlation between musical ability (musical memory and pitch discrimination) and the ability to recognize Mandarin tonemes?
2. If this correlation between musical ability and phonological perception exists, does it differ by gender or age?

METHODOLOGY

Participants watched a video introducing the Chinese tonemes. Participants then completed two assessments related to Mandarin tonemes and one related to musical ability. In an online survey, participants reported their scores and demographic information. A Pearson's correlation was calculated between the scores on the musical ability and phonological perception assessments.

Participant tasks

Participants watched an instructional video, completed 3 assessments, and completed a survey.

Instructional video. Participants viewed a video that explains Mandarin tones in beginners' terms. The video uses self-explanatory terms ("falling," "rising," "flat," "falling and then rising") instead of "tone 1, "tone 2," etc. This terminology helped to ensure that the test measured participants' recognition of tones, not their ability to remember arbitrary labels.

The video appeals to different learning/sensory preferences. The instructor uses hand gestures to indicate pitch contours (appealing to kinesthetic learners), color coding, pictures, and marks over vowels (appealing to visual learners), and spoken demonstrations of the tones themselves (appealing to auditory learners).

Musical memory and pitch perception test. Dr. Jake Mandell, M.D., developed the pitch perception test for research on neuro-anatomical correlates of congenital amusia, or "tone deafness." Mandell used the test to measure his patients' pitch discrimination and musical memory abilities, both key factors in musical ability. People with amusia score very low on assessments of both of these factors, while professional musicians typically score very high. In this study, the participants' task was to listen to 36 pairs of melodies and indicate whether the two melodies are the same or different. The test returns a score that represents the percentage of correct answers.

Mandarin tone games. The Mandarin tone games are part of the Chinese language-learning website developed by the British Broadcasting Corporation (BBC). The games were designed as a review for Mandarin learners. The game consists of 3 multiple-choice tests: Easy (ten 1-syllable words), Medium (ten 2-syllable words), and Hard (ten 3-syllable words). Players listen to a word and choose the correct pinyin representation of the word (example: má, mà, mā, mǎ). Participants first completed the Easy game in order to acquaint themselves with the process and user interface. Then they completed the Medium game, whose data was used for correlation with the musical assessments.

Participants reported their assessment scores and demographic data in an online survey. The survey asked for their scores on the musical ability assessment, scores for both of the Mandarin tone games, as well as age, gender, native language, other languages spoken and studied, and whether participants were linguistics majors (linguists were excluded from the study). A comments field allowed participants to make notations about any extenuating circumstances (hearing loss, tinnitus, problems accessing the assessments, etc.).

Participants. The survey was open only to adults with no tonal language background. Participants were recruited through convenience sampling, using social media and personal contacts. After the removal of "disqualified" responses (linguists, incomplete surveys, and obvious score errors), data were used from 107 participants, with the following demographic breakdown.

Table 1

Number of participants by gender

Gender	Number of participants
Men	48
Women	58
Other/No answer	1

The ages of participants ranged from 18 to 80, with a median age of 45.

Table 2

Number of participants by age range

Age range	Number of participants
18–25	22
26–35	15
36–45	17
46–55	18
56–65	25
66–80	10

RESULTS

Assessment scores

Scores on the musical ability assessment ranged from 26.1 to 100 (on a 100-point scale), with a mean of 74.5 and a median of 75.0. Scores on the medium Mandarin tone game ranged from 0 to 10 (on a 10-point scale), with a mean of 4.8 and a median of 5.

Table 3

Assessment scores

Assessment	Lowest score	Highest score	Mean	Median
Musical ability	26.1	100	74.5	75
Mandarin tones	0	10	4.8	5

Correlations

The correlation between the musical test scores and the Mandarin tone game for all participants was .340.

Gender

The correlation for men (.353) was only slightly higher than for women (.334), not constituting a statistically significant difference, and their average assessment scores were also comparable.

Table 4

Musical ability and toneme score correlations by gender

Gender	Correlation
Men	.353
Women	.334
Overall	.340

Table 5

Average scores by gender

Gender	Number of participants	Average scores	
		Musical ability	Toneme recognition
Men	48	75.8	5.1
Women	58	74.3	4.6
Other	1		
Overall	107	74.5	4.8

Age

The correlation varied by age, with the older participants showing the highest correlation. The median age of participants was 45.

Table 6

Musical ability and toneme recognition score correlations by age (by median)

Age	Correlation
18-44	.278
45-80	.396
Overall	.340

The average test scores for the two groups were not statistically different, although the younger half of the participants scored slightly higher on both tests:

Table 7

Average test scores by age (by median)

Age range	Average test scores	
	Musical ability	Toneme recognition
18-44	75.5	5.0
45-80	73.6	4.6
Overall	74.5	4.8

Breaking the age ranges down into smaller groups shows that the oldest participants (ages 66 through 80) had the highest correlation between their musical and toneme recognition scores. However, given the smaller size of this portion of the participant sample, it is possible that the correlation for this group is not representative of the overall population.

Table 8

Musical ability and toneme recognition score correlations by age

Age range	Number of participants	Correlation
18-25	22	.348
26-35	15	.057
36-45	17	.469
46-55	18	.331
56-65	25	.287
66-80	10	.599
Overall	107	.340

While the oldest participants had the highest correlation between their scores, they also had the lowest scores on both assessments.

Table 9

Average scores by age

Age range	Number of participants	Average scores	
		Musical ability	Toneme recognition
18-25	22	73.0	5.2
26-35	15	78.3	5.1
36-45	17	75.7	4.7
46-55	18	73.0	4.3
56-65	25	76.0	5.2
66-80	10	69.3	3.9
Overall	107	74.5	4.8

Summary

As the correlations listed above demonstrate, there is a positive correlation (.340) between scores on musical ability tests and Mandarin toneme recognition assessments. The correlation is very slightly higher for men (.353) than for women (.334) and somewhat higher for participants who are older than the median age (.396) than for younger participants (.278).

DISCUSSION

This study reinforces findings by previous researchers, provides new insight into phonological perception variation across adult age groups, and suggests a need for additional research on the effects of age and gender on adult language learners.

The current study, because it relied on convenience sampling to recruit participants, did not have as homogeneous a participant population as the Nakata study, whose adult participants included only those without formal music training; or the Slevc and Miyake study, which included only Japanese-L1 participants; or the Bowles, Chang, and Karuzis study, whose participants were all young adults. The weaker correlation in the current study might suggest that the participants' linguistic diversity was a confounding variable that dampened the correlation.

The current study did, however, begin to explore whether the correlation between musical ability and language-learning ability differ by gender or by age. While no significant difference in correlation was found by gender, the current study suggests that the link between musical ability and language-learning ability may change over the adult lifespan. The differences in correlation did not exhibit a consistent change from decade to decade, but the differences in correlation between the participants above and below the median age suggest that more research, with participants who are more homogeneous in factors other than age (for example, with the same language background), is needed to determine more clearly how age affects the correlation between musical ability and language-learning ability.

Further research

The discrepancies between groups within this study's participant population—most notably by age range—suggest a need for further study of the effects of age differences in language learning. While a great deal of research has been done comparing children and adults in terms of their language-learning outcomes, little research has been done to compare language learning between younger adults, middle-aged adults, and older adults. Further studies would need to control for general cognitive factors to isolate the changes in language-learning ability, independent of potential decline in general cognitive abilities. With the growing popularity of language-learning apps, such research could inform the design of better language-learning tools for adults.

CONCLUSION

This experiment demonstrated that short-term musical memory and pitch perception correlated positively with the ability to distinguish between Mandarin tonemes, providing some support for the belief that there is a positive relationship between musical ability and language learning.

However, a wide range of individual differences, and complex interactions between these individual differences, also have an influence on language-learning achievement, and it is important not to oversimplify the effect of any one influence in isolation from the others.

While the study of these other factors was beyond the scope of this project, it would be useful to find a correlation that could be used, even cautiously, to predict success in specific aspects of second-language learning, in this case, phonological perception. Such predictions could influence language learning strategies and teaching methods to benefit from learners' musical ability or compensate for a lack thereof, based on pretests. Extra exercises might be designed to help students develop their perception skills.

Previous research indicated a need for more research into the music-language link. While this study provides insights into the effects of age on this link, more research is needed to understand how age influences language learning. Because both musical ability and language-learning ability comprise several specific sub-skills, it remains to be established which of these musical sub-skills correlate with specific language sub-skills.

ABOUT THE AUTHOR

Jane Lorenzen is a graduate student at the University of North Texas. Her interests are individual differences in adult language learning, endangered language revitalization, and eye-tracking applications in linguistics research.

Contact information:
JaneLorenzen@my.unt.edu
P.O.B. 261774
Plano TX 75026
972.379.7867

REFERENCES

BBC (n.d.) *Chinese game—Tone game*. Retrieved from www.bbc.co.uk/languages/chinese/games/tones.shtml

Bowles, A. R., Chang, C. B., & Karuzis, V. P. (2016). Pitch ability as an aptitude for tone learning. *Language Learning*, *66*(4), 774–808. doi:10.1111/lang.12159

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In Y. Miyashita, A. Marantz, & W. O'Neil (Eds.), *Image*, *language*, *brain*, (pp. 95–126). Cambridge, MA: MIT Press.

Koelsch, S., Gunter, T. P., von Cramon, D. Y., Zysset, S., Lohman, G., & Friederici, A. D. (2002). Bach speaks: A cortical 'language-network' serves the processing of music. *NeuroImage*, *17*(2), 956–966. doi:10.1006/nimg.2002.1154

Mandell, J. (n.d.) *Tonedeaf Test:* Retrieved from www.jakemandell.com/tonedeaf/

McDowell, K. D. (2004). *Examining relations among expressive phonology, phonological processing, and early decoding skills in children* (Doctoral dissertation). Retrieved from Florida State University Library, *Electronic Theses, Treatises and Dissertations*, paper 2572.

Milovanov, R., & Tervaniemi, M. (2011). The interplay between musical and linguistic aptitudes: A review. *Frontiers in Psychology*, *2*, 1–6. doi:10.3389/fpsyg.2011.00321

Nakata, H. (2002). Correlations between musical and Japanese phonetic aptitudes by native speakers of English. *Reading Working Papers in Linguistics, 6*, 1–23.

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, *6*, 674–681. doi:10.1038/nn1082

Peynircioğlu, Z. F., Durgunoğlu, A. Y., & Öney-Küsefoğlu, B. (2002). Phonological awareness and musical aptitude. *Journal of Research in Reading*, *2*5(1) 68–80. doi:10.1111/1467-9817.00159

Skehan, P. (1989). *Individual differences in second language learning*. New York: Routledge.

Slevc, L. R., & Miyake, A. (2006). Individual differences in second-language proficiency: Does musical ability matter? *Psychological Science*, *17*(8), 675–681. doi: 10.1111/j.1467-9280.2006.01765.

Tian, F. [Mandarin MadeEZ by ChinesePod]. (2012 July 20). *Tones in Mandarin Chinese with Fiona Tian: Lesson 1* [Video file]. Retrieved from www.youtube.com/watch?v=4ZB9-vZFesk

Tillman, B., Janata, P., & Bharucha, J. J. (2003). Activation of the inferior frontal cortex in musical priming. *Cognitive Brain Research*, *16*(2), 145–161. doi:10.1016/S0926-6410(02)00245-8

Nowacka, M. (2019). Progress testing after two-semester pronunciation instruction: Spelling-pronunciation. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Ames, IA, September 2018 (pp. 318-340). Ames, IA: Iowa State University.

PRESENTATION/POSTER

PROGRESS TESTING AFTER TWO-SEMESTER PRONUNCIATION INSTRUCTION: SPELLING-PRONUNCIATION

Marta Nowacka, University of Rzeszów, Poland

The primary aim of this study is to determine whether English Department students' pronunciation progressed during a one-year course of practical and theoretical phonetic instruction and, if so, to verify in what respects. A second intention was to discover what problems still remain despite the course. A self-designed diagnostic test was administered to 91 first-year students at the beginning (pre-test) and at the end of the course (post-test). The word-reading exercise encompassed 35 lexemes (43 aspects) that exhibited a variety of difficulties, including problematic letters, e.g. <o> in *oven* versus *protein*, <ch> in *charlatan* versus *archives* and words commonly mispronounced (*ancient*) together with examples showing frequent word-stress misplacement (*purchase*). The sentence-reading task (30 elements) comprised: weak forms, contractions (*mustn't*), a selection of 'trap' words (*dough*), words with difficult word stress (*determined*) and rendition of verb forms. This evidence-based testing method suggests that the one-year course is beneficial because it leads to the participants making overall progress (r = 0.71 for word-reading, r = 0.75 for sentence reading in pre-test and post-test). It also shows that contracted forms and some phonetically challenging words (*area, purchase, Niagara*) still call for attention.

INTRODUCTION

In the contemporary multidimensional approach to phonetic teaching, accent is interrelated with comprehensibility, intelligibility and fluency in communication. Spelling-pronunciation, inappropriate inference from orthography, has been found to have a negative effect on what the interlocutor understands and on ease of decoding a message. Wells (2008) provides a solution to spelling-pronunciation arguing that, "either we must reform English spelling […] or teachers of English to speakers of other languages must *teach the pronunciation* of each word as well as its spelling" (p. 104).

Dickerson (2015) believes that spelling is a valuable resource for English as a Second Language (ESL) and English as a Foreign Language (EFL) learners, working for the benefit of their oral accuracy and fluency. He stresses the fact that the use of orthography serves prediction most directly and, through making good judgments, it serves perception and production. He indicates that orthography can be implemented for predicting the following: consonants, major word stress, major stressed vowels, compression, suffix forms and variability. He remarks that by giving students access to some orthographic rules we provide them with life-long knowledge of sound via spelling.

In a substantial number of recent studies on Polish-accented English, spelling-pronunciation is recurrently salient in the hierarchy of errors (Bryła-Cruz, 2016; Nowacka, 2016, 2018b; Porzuczek, 2015; Szpyra-Kozłowska 2013, 2015; Zając, 2015). This research shows that spelling-induced pronunciation errors and whole words with deceptive spelling hamper intelligibility and thus constitute one of the priorities in teaching pronunciation. For example, Bryła-Cruz's (2016) comprehensive research into the perception of Polish-accented English established a list of pronunciation priorities. These priorities include eliminating spelling-based errors followed by the dental fricatives, velar nasal, vocalic contrasts (STRUT vs. BATH vs. TRAP, FLEECE vs. KIT, NURSE vs. DRESS, NORTH vs. GOAT), word stress, maintaining voicing of lenis obstruents and weak forms. Because spelling-induced pronunciation errors proved critical to the four examined parameters included in Bryła-Cruz's study (i.e, accentedness, comprehensibility, intelligibility and irritation), the present study investigates the claim for the need to interrelate orthography with pronunciation during phonetic training. It also attempts to show that an explicit focus on phonetically challenging words and the inclusion of some orthographic rules in the phonetic training leads to the eradication of some spelling-induced pronunciation errors.

Research questions

This paper on summative assessment aims to determine phonetic attainment after one-year of pronunciation instruction. It intends to firstly provide evidence about the first-year English Department students' progress by comparing their initial and final performance of 73 phonetic items in word reading and sentence-reading and secondly, to adjust a syllabus and design materials to cater to learners' pronunciation needs. It seeks to answer the following questions:

1. Is there any progress in learners' pronunciation of 73 aspects in 61 phonetically challenging lexemes (and/or words with deceptive spelling) after the two-semester pronunciation instruction?
2. Which words and phonetic aspects have been learnt?
3. What phonetic problems still remain after the course?

METHODS

Participants

The research concerns a specific group of foreign language learners, 91 first-year students of English, at the University of Rzeszow, Poland, who were taught English phonetics and pronunciation by the author. 59% (*n*=54) were full-time and 41% (*n*=37) were part-time students of which females constituted 73% (*n*=66) and males 27% (*n*=25). Their age ranged from 19 to 36 years. Most students, 82% (n=73), were between the ages of 19 and 21. The mean length of compulsory institutional FL learning is 12 years. Most students report having learnt English for about 15 years.

The study participants were preparing to become English teachers and/or interpreters or were likely to work in a linguistic environment at schools, universities, etc. They studied phonetics only during the first year of their university studies. The total number of hours of phonetics the students receive at university depends on the type of the course and ranges from 40 hours (20 hours of English phonetics lectures and 20 hours of practical pronunciation classes) for part-time students to 90 hours (30 hours of English phonetics lectures and 60 hours of practical pronunciation classes) for full-time students. The lectures cover fundamental topics in phonetics (e.g., basic terminology, articulators, production of speech, transcription) in both segmental (vowels, consonants) as well as suprasegmental phonetics (e.g. word stress, weak forms, rhythm, linking, elision and assimilation), while the practical course corresponds to the content of the lectures.

The course was delivered by one teacher, the author herself, which guaranteed that the participants received the same quality of instruction. In regards a typical study procedure from the very beginning, apart from regular work on English segments and suprasegments, during every lesson, 5 to 10 minutes were devoted to the explicit teaching of the relationships between spelling and pronunciation in the form of rules, concerning such issues as regular inflections in suffix forms, such as past tense/past participle *–ed*, and most typical letter-to-sound correspondences concerning a vowel or a consonant. Three to four complete lessons focused on words with deceptive spelling, the list of over 600 *Words Commonly Mispronounced* (Sobkowiak, 1996) and the relationship between orthography and pronunciation, such as summarized spelling guidelines (Collins & Mees, 2008).

Materials, test design and instructional items

The study implemented a diagnostic pronunciation test designed by the author which included two reading aloud tasks (see Appendix 1) to target a particular phonetic feature for evaluation.

The 61 lexical items, which included 73 selected phonetic targets, were covered during the one-year English phonetics instruction. They come from a variety of teaching materials, which supplement the main coursebook by Roach (2009) and a workbook by Mańkowska et al. (2009). In brief, these resources encompass: transcription of irregular verbs (Sobkowiak, 1996) and the above-mentioned commonly mispronounced words; transcription and awareness-raising exercises on deceptive spelling and challenging words (Sobkowiak & Szpyra, 2001); spelling guidelines (Collins & Mees, 2008); Nolst Trenité's poem *Chaos*, known for the inconsistencies of English spelling (Upward, 1994); contracted forms (Lewis, n.d.); strong and weak forms (Lecumberri & Maidment, 2000).

The selected features in word reading regard segments and silent consonantal letters, stress placement and the suffix *–ate*. Among the short vowels there were the vowels DRESS [i] (*sweat*, *threat*), TRAP (*chassis*), STRUT (*oven*, *sponge*), LOT (*foreign*, *cough*) and commA (*thorough*, *charlatan* - unstressed syllables). Long vowels were the vowels FLEECE (*protein*, *fiend*, *suite*), NURSE (*word*, *purchase*, *courteous*), GOOSE (*ghoul*, *feud*), THOUGHT, /ɑː/ in American English, (*author*, *gnaw*, *hawk*, *saw*, *abroad*) and BATH/START (*draught*, *sergeant*). The analysis also comprises four diphthongs: the vowels in FACE (*ancient*, *failure*, *steak*), PRICE (*disciple*), GOAT (*protein*, *comb*, *folk*) and SQUARE (*area* and *scarce*). The consonants focused on the pronunciation of a diagraph <ch> (as /ʃ/ in *charlatan*, /ʃ/ (Br) or /tʃ/ (Am) in *chassis* and /k/ in *archives*) and silent consonantal letter(s) such as <b> in *comb*, <l> in *folk* and <g> of initial <gn> in *gnaw*. Stress placement was examined in *foreign*, *protein* and *purchase* and in *accurate*, *area*, *chassis*, *disciple* and *satire*. An unstressed suffix *–ate* (/-ət/, /-ɪt/) was tested in *accurate*. The pronunciation of the 'whole words' encompasses stressed vowels and: unstressed syllables, e.g. in *ancient*, *area*, *chassis*, *courteous*, *disciple*, *failure*, *foreign*, *purchase*, *thorough*, or a set of aspects, e.g. in *archives* (START, /k/, PRICE), *draught* (final consonants), *satire* (TRAP, triphthong: PRICE + COMMA) and *sergeant* (<ge>≡/dʒ/).

In the classification of phonetic elements in sentence-reading there are contractions, weak/strong forms and content words. Contracted forms are controlled for: GOAT in *don't, won't*, STRUT and /s/ in *mustn't* and NORTH/CURE in *you're*. The weak form list comprises: *has*, *have*, *of*, *that*, *the* (as /ði/ before a vowel), *would* and one strong form *some*. The content words include: adjectives (*appalled*, *available*, *basic*, *determined*, *nauseous*), nouns (*chaos*, *course*, *dough*, *lager*, *leopard*, *Niagara*, *pint*, *prayer*, *pronunciation*, *yolk*) and verbs (*develop*, *draw*, *lay*, *risen*). In all monosyllabic and two disyllabic words the analysis is restricted to a vowel, e.g. NORTH in *course*, GOAT in *dough*, *yolk*, PRICE in *pint*, THOUGHT or PALM in *draw*, FACE in *lay*, SQUARE or DRESS/TRAP in *prayer* - referring to words of praying, and KIT in *risen*. One group of polysyllabic words focuses on at least two features such as vocalic and consonantal sound or a letter-to-sound correspondence, e.g. FACE and /s/ in *basic*; FACE, LOT or PALM and <ch> as /k/ in *chaos* and PALM and <g> as /g/ in *lager*. Yet another set of polysyllabic words examined stress placement, the quality of a stressed vowel and, except *appalled* (THOUGHT) and *determined* (NURSE), unstressed syllables, e.g. FACE in *available*, THOUGHT or PALM in *nauseous*, DRESS in *leopard*, TRAP in *Niagara*, STRUT in *pronunciation* and DRESS in *develop*. For reasons of brevity in the discussion of the results whole words rather than the specific phonetic feature or features examined in them are referred to.

Appendix 2 presents a detailed examination of each lexical item with the name of the teaching material it is taken from, referred to as a source, the description of an examined aspect or aspects and an example of an error as well as of accepted pronunciation.

Procedures

There were two stages of the data collection: a pre-test recorded in October 2017 in the first week of students' study encompassing task 1 (word reading: 43 items) and task 2 (sentence-reading: 30 items); and a post-test gathered in May 2018, in the final weeks of the second term. An evaluation of the respondents' renditions of words used the following protocol. The students were asked to produce the words and utterances in the diagnostic test in the way they found easiest to pronounce. We did not insist on a single pronunciation of a word but found it justifiable to accept the educated standard variants of British and General American English, the most frequently learnt varieties of English by Poles. Other inner-circle varieties of English were not observed to have been applied by the students in this research. The recordings were then evaluated over the period of two months by one rater, the author of the text, with a PhD in linguistics and over 20-years' experience in teaching and researching pronunciation of Polish students of the English language. Each student's speech was transcribed and then on the basis of a rendition of a word, a point or zero was assigned to a student for their enunciation of an examined aspect in a word, e.g. stress placement in *characterize* - 1 point for /ˈkærəktəraɪz/, zero for ≠ /kəˈræktəraɪz/ or ≠ /kəˌræktəˈraɪz/. Then the following statistical tests were applied: Wilcoxon signed-rank test for correlation coefficient between the number of points in the pre-test and post-test and Cochran's Q test to check the statistical significance of the pre-test and post-test results. Both quantitative and qualitative data were gathered on the basis of this assessment. This paper focuses on the quantitative part and uses the qualitative data from transcriptions of errors to clarify the nature of the problems.

RESULTS

General progress

The answer to the first research question, which examined if there was any progress in students' pronunciation of 73 phonetic aspects in 61 lexical items after the one-year instruction, is positive. In both tasks, word reading and sentence reading, the coefficient of 0.71 and 0.75 respectively indicates a positive and directly proportional correlation.

The Wilcoxon signed-rank test was applied, and it revealed that in both tasks the difference between the two tests, pre-test and post-test, is statistically significant $p<\alpha$ ($p=0.00000$). This paired difference test is a non-parametric statistical hypothesis test used to compare two matched measurements on a single sample to assess whether their population mean ranks differ. It can be used as an alternative to the paired Student's t-test when the population cannot be assumed to be normally distributed.

Figure 1 shows the mean number of points the respondents scored in Task 1 and 2 in pre- and post-tests. In word-reading the initial mean 17.2 grew to final 26.6 out of the maximum 43. The initial standard deviation 6.9 increased eventually to 7.8 while the pre-test minimum and maximum 3 and 38 moved up to a post-test of 5 to 41. In sentence-reading the mean rose from 13.4 (pre-test) to 19.5 (post-test) out of the maximum 30. The beginning standard deviation of 5.1 increased to 6.0 while the pre-test minimum of 5 stayed the same and the maximum of 29 rose to 30. A higher standard deviation in both post-tests shows that there is a greater differentiation. In other words, correct renditions are spread over a wider range of values.

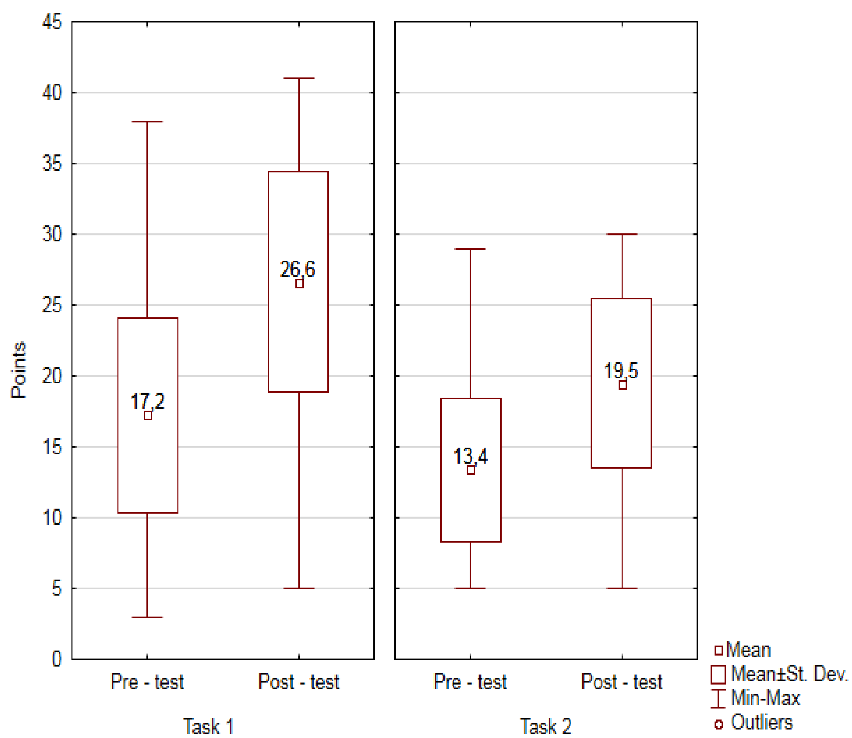


Figure 1. The mean number of correct renditions in pre-test and post-test in Tasks 1 and 2.

Detailed progress: Words and phonetic aspects

Research question two asked which words and phonetic aspects had been learnt. The Cochran Q test was applied to determine if there had been a change in the pronunciation of a given phonetic feature in a pre-test and post-test. In this non-parametric statistical test with a binary response, the variable takes only two possible outcomes, coded as 0 for failure and 1 for success.

The test revealed that p was less than α (α=0.05), thus a significant change was observed, for most words in Task 1, except for the following: *chassis*_<ch> (0.057), *draught* (0.057), *suite* (0.117), *charlatan*_COMMA (0.126), *scarce* (0.317), *protein*_S (0.423), *fiend* (0.601), *area* (0.705) and *satire* (0.808).

Figure 2 exemplifies that in word reading the highest rate of progress between pre-test and post-test was noted for: *comb <b>* (80%), *sponge* (44%), *abroad* (43%), *author* (40%), *gnaw <gn>* (36%), *foreign_W* (34%); *failure*, *folk <l>*, *purchase_W*, *thorough*, *word* (32% each); *gnaw_V*, *saw*, *steak* (29% each); *accurate*, *hawk* (27% each); *folk*_V, *oven* and *protein_V* (26% each). The most striking improvement (i.e. of 80%) was noticed for the enunciation of *comb <b>*, which might be due to awareness that was developed from the explicit training, learning a rule of a silent letter <b> in a final letter combination <mb>, and the insertion of GOAT vowel. All cases of progress over 8%, with the exception of *suite* (10%) are statistically significant, thus the non-significant differences, marked in grey, include: *suite* (10%), *charlatan_COMMA* (8%), *chassis <ch>* (7%), *scarce* (7%), *draught* (7%), *protein_S* (4%), *fiend* (3%), *satire* (-1%) and *area* (-2%).

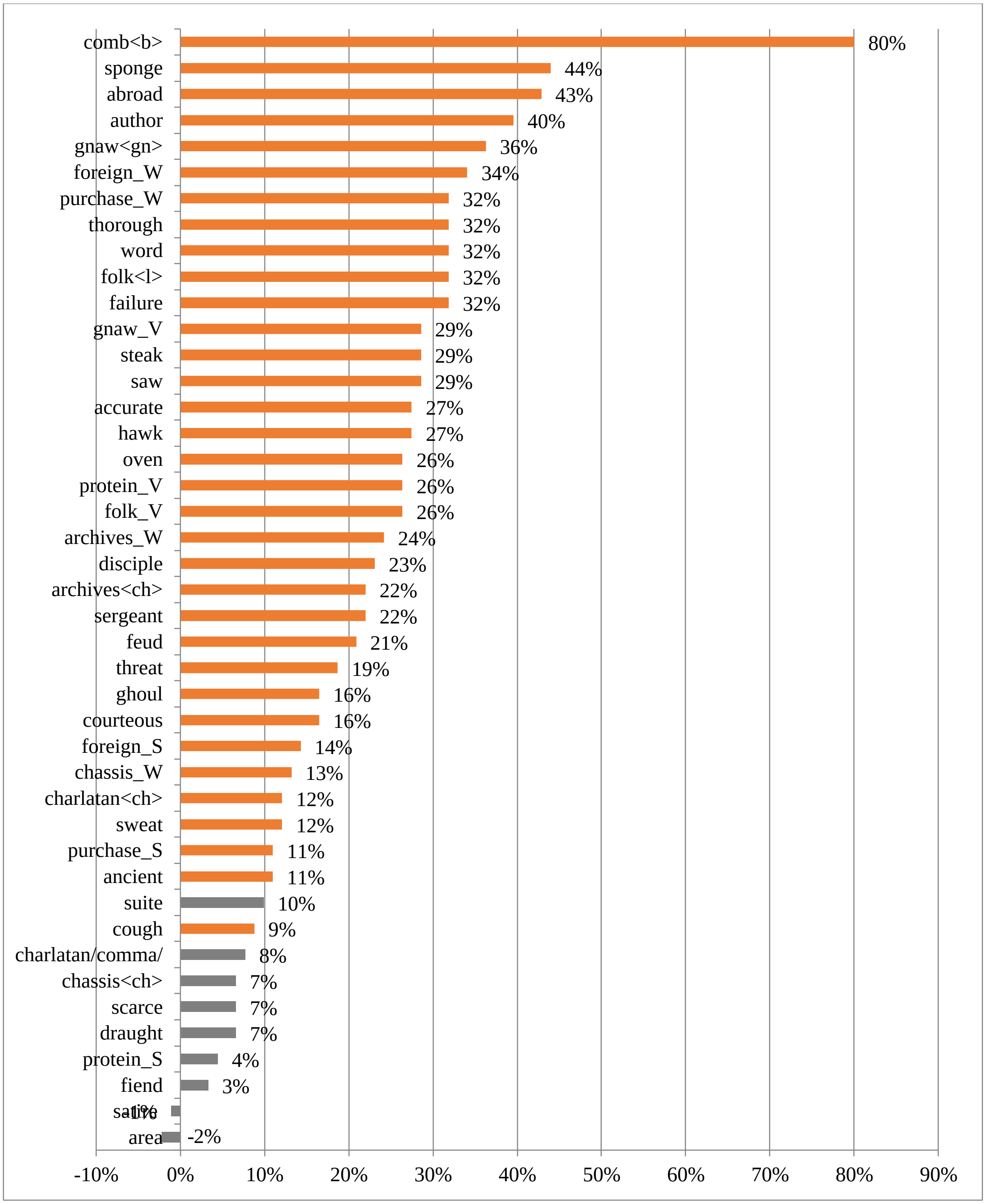


Figure 2. The percentage of progress between post-test and pre-test results in word reading.

In Task 2 the Cochran Q test revealed that p was less than α (α=0.05), which means that the overall progress was significant in all words except *you're* (0.105), *would_l* (0.131), *don't* (0.165), *lay* (0.256), *basic* (0.512), *course* (0.831) and *mustn't* (1.00).

Figure 3 shows the rate of progress for phonetic aspects in sentence-reading. Three top words improved by more than 40%: *leopard* (45%), the two weak forms *have* (42%) in *I must have lost it* and *of* (41%) in *part of an egg*. Then, there is *pronunciation* (29%), a strong form of *some* (26%), a weak-form conjunction *that* (26%), *draw* and *determined* - both with 25%.

Other phonetic elements improved by less than 25%, except for *mustn't* which stayed the same and *course* (-1%) which regressed slightly. The progress rate in the following words ranges from 8% to 24%. Included here are *nauseous* (24%), *lager* (24%), *chaos* (23%), *have* in *I have ever seen* (22%), *of* in *the apple of my eye* (22%), *prayer* (19%) as 'words of praying', *risen* (19%), *yolk_V* (18%), *the* in *the apple* (18%), *dough* (18%), *pint* (18%), *won't* (18%), *Niagara* (16%), *available* (14%), *has* as in *The sun has just risen* (13%), *develop* (12%), *appalled* (11%), *would* in *she would come* (10%) and *the* in *the end* (8%). In only 7 cases, marked in grey at the bottom of the graph with the lowest scores, the difference between the final and initial pronunciation was of no statistical significance, except *the_e*.

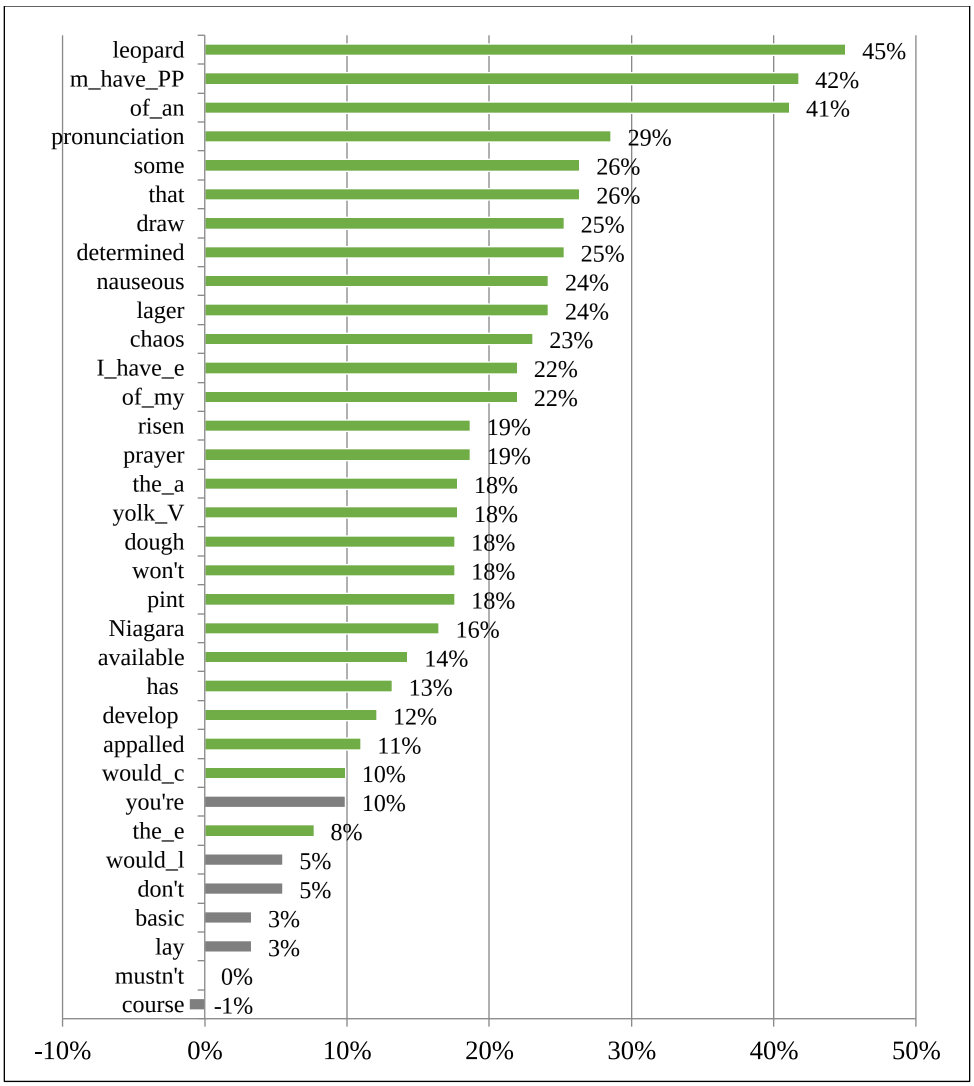


Figure 3. The percentage of progress between post-test and pre-test results in sentence-reading.

Word frequencies of occurrence as well as their rank orders[ii] were examined in iWeb, the largest existing corpus of the English language that is about 14 billion words in size, to see if they had affected the obtained results. The words selected for the analysis belong to the top 60,000 lemmas in the iWeb corpus. There were only eight medium frequency words (~ #25000 rank[iii]) such as *saw*, *fiend*, *draught*, *ghoul*, *nauseous*, *gnaw*, *appalled* and *charlatan*. This part of the study does not comprise low frequency words, which are around rank #45,000. In brief, having juxtaposed the word frequency and a rate of progress shows there is no straightforward linear correlation between the two variables, e.g. *comb* (#9454) improved by 80%, *sponge* (#7563) got better by 44% while *area* (#153) deteriorated by 2% and *satire* (#13877) by 1%. A detailed examination of these results is presented in an unabridged version of this study (Nowacka, in progress).

Progress in broad phonetic categories: Pre-test and post-test across two tasks

Having grouped individual phonetics items into five broader categories, which had been studied during the course, some improvement was observed in all cases between initial and final pronunciation. In word reading (Figure 4) the greatest progress of 25% is found in the rendition of ‘words commonly mispronounced’, followed by a substantial 19% increase in spelling-conditioned pronunciation (e.g. in *gnaw*, *archives*), that is, one’s familiarity with an orthographic rule, regarding for example the silent letter <b> in a final <mb> sequence of letters as in *comb* and *bomb*. Words from the poem *Chaos* (e.g. *chaos*, *disciple*) improved by 16% and in other challenging words outside the previously discussed groups, such as *dough* and *Niagara*, an 11% increase in correctness was noted.

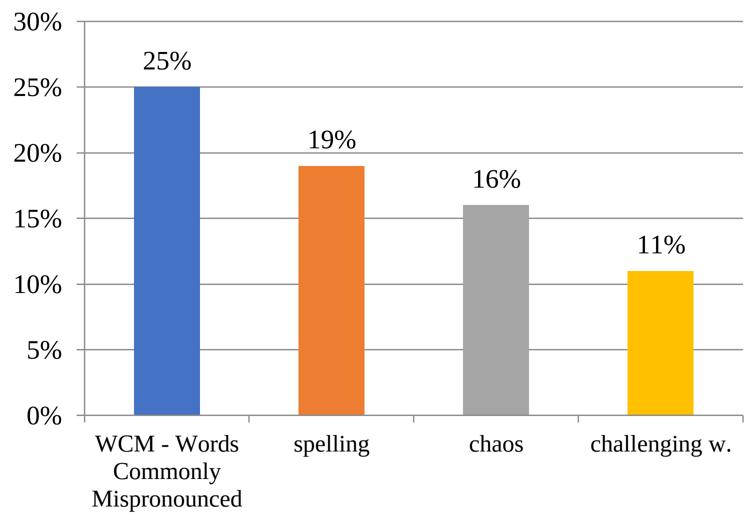


Figure 4. Cumulative progress for broader categories in word reading (WR).

In Figure 5, referring to sentence-reading, there is a 24% improvement with ‘Chaos’ and ‘other challenging words’. The correct rendition of weak and strong forms rises by 21%, verbs progress by 16%, ‘words commonly mispronounced’ by 12% and contracted forms by 8%.

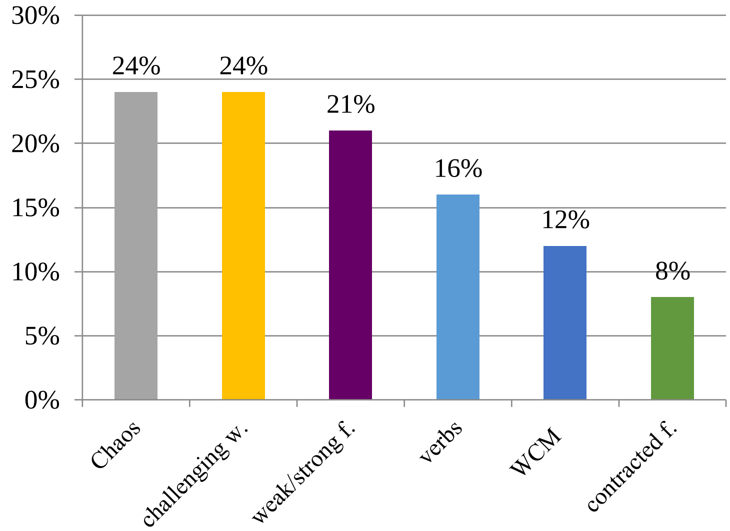


Figure 5. Cumulative progress for broader categories in sentence-reading (SR).

The results for cumulative phonetic categories across the two tasks presented in Figures 4 and 5 reveal that depending on the type of task and/or a selection of words the degree of progress varies.

Remaining phonetic issues

To answer research question 3 about phonetic problems that remain in spite of phonetic training, first we examined under 50% results in Figure 6 corresponding to word reading. In this group of words that are mispronounced by a majority of the informants, there are eight phonetically challenging words: *chassis_W* (18%), *courteous* (21%), *ancient* (25%), *sergeant* (27%), *thorough* (34%), *disciple* (35%), *archives_W* (41%), *feud* (44%); 2 problematic letter-to-sound correspondences regarding <o>, i.e. representing GOAT in *folk_V* (46%) and STRUT in *oven* (47%); and five more less frequent words with non-significant progress, such as *draught* (10%), *satire* (12%), *protein* (39%), *suite* (44%) and *scarce* (45%).

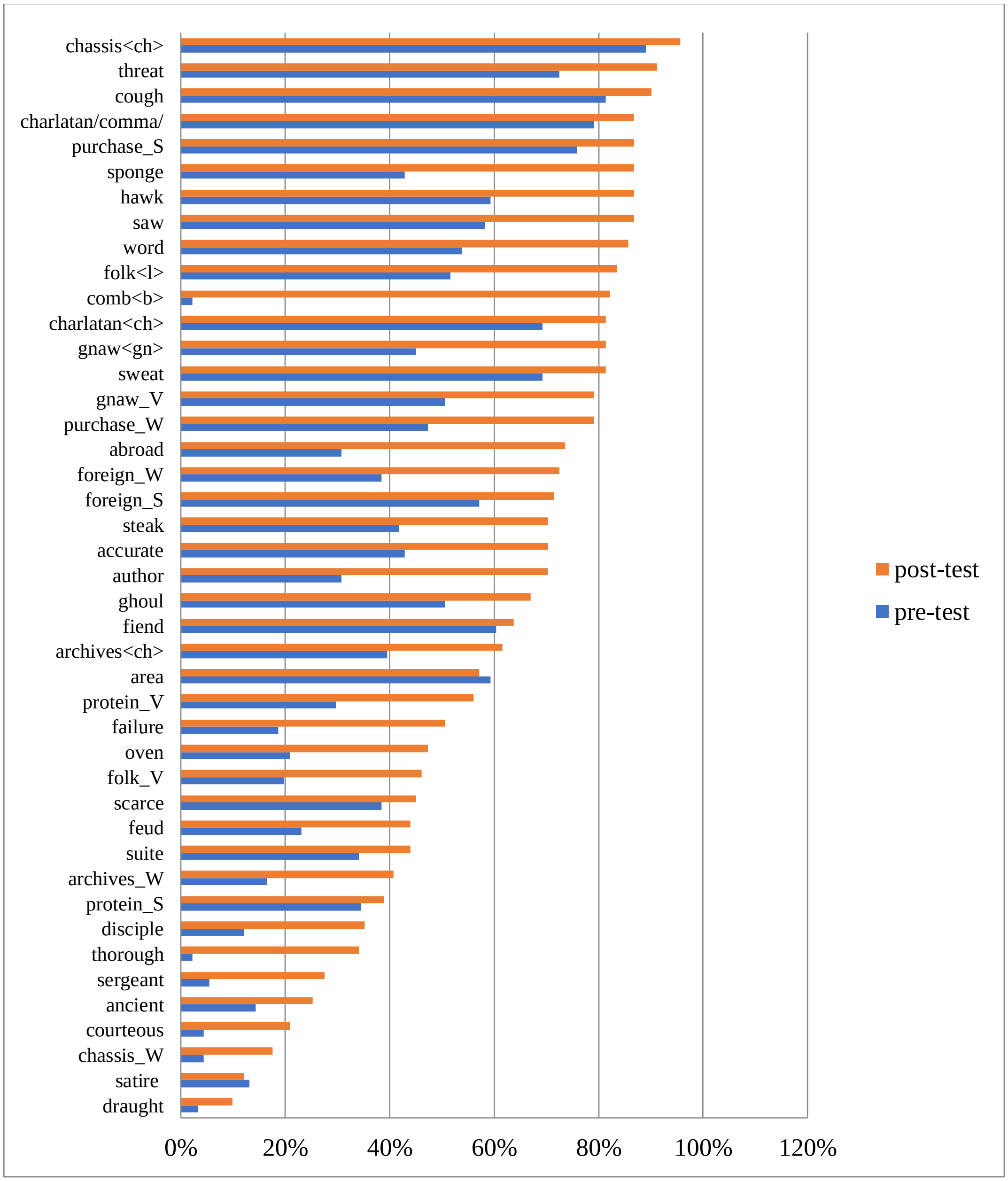


Figure 6. The pre-test and post-test results in word reading.

Figure 7 on sentence-reading reveals that the following items were not completely learnt by the majority of students. These include eight phonetically difficult content words, *prayer* (21%), *Niagara* (24%), *pint* (24%), *yolk* (29%), *developed* (38%), *appalled* (40%), *nauseous* (47%), *risen* (49%), four contractions, *don't* (14%), *won't* (26%), *mustn't* (36%), *you're* (49%), and two weak forms, *has* (25%), *of* (35%).

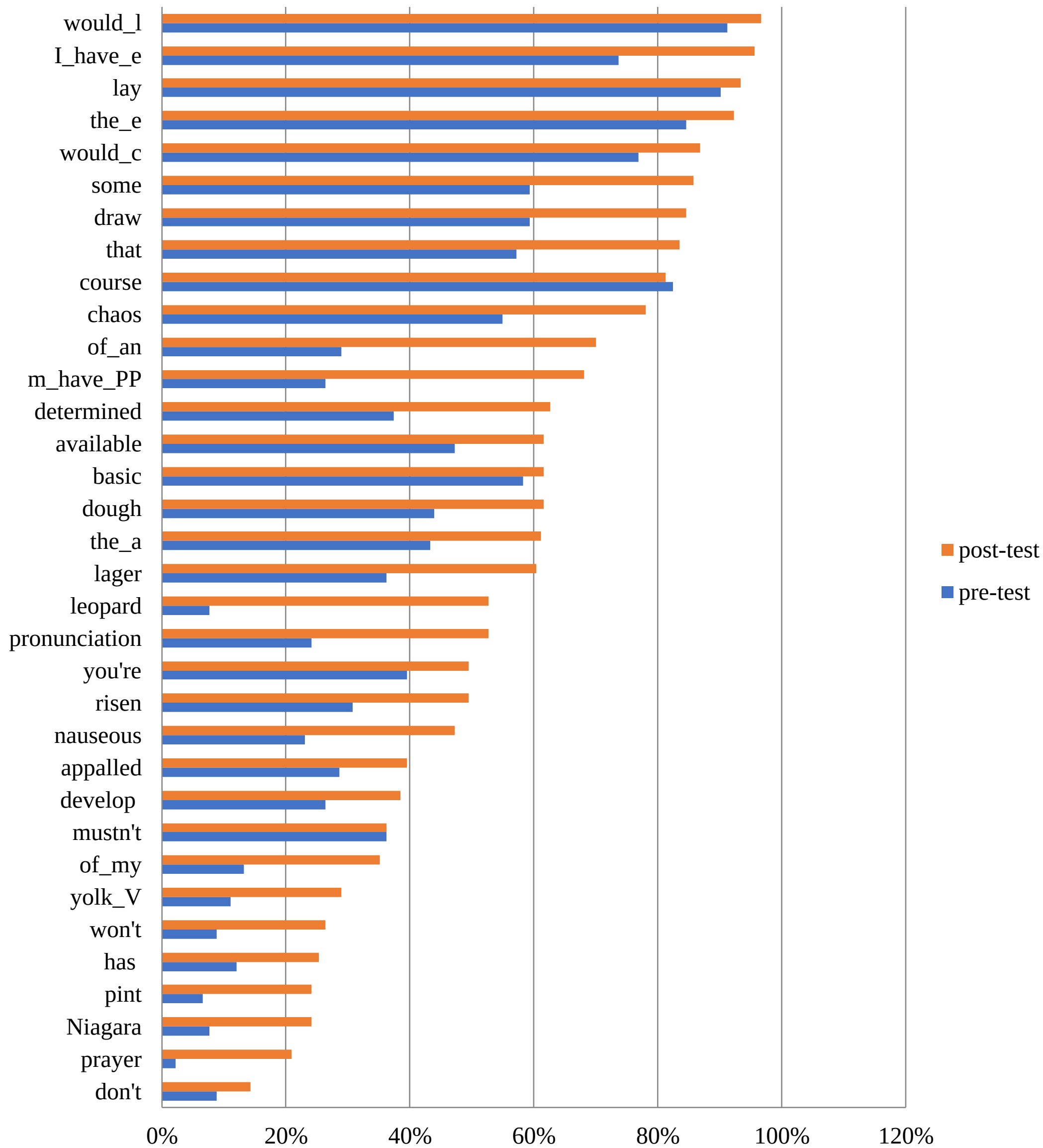


Figure 7. The pre-test and post-test results in sentence-reading.

DISCUSSION

There are several limitations to this study. The single rater assessment is a shortcoming as it does not allow for inter-rater reliability to be calculated. The two-month evaluation period was intended to eliminate the effect of fatigue on the rater's assessment, however, it was not able to remove possible rater subjectivity or inconsistency. It is hoped that the representative sample of nearly 100 respondents might balance this imperfection. Another limitation is the lack of a control group. This research was meant as a progress test of a specific group of students undergoing phonetic instructions, thus it would be highly unethical not to teach pronunciation to the English Department students to have access to a research control group. Even without it, the scope of the research, with 73 phonetic elements scrutinized, offers an abundance to learn from.

The next step in this research is the analysis of qualitative data to shed light on the range of renditions of the tested material. This knowledge could be applied while preparing a multiple-choice task in a written test of pronunciation.

What needs to be checked is the progress of full-time and part-time students in search of similarities and differences that should also be reflected in the syllabus. The likely dissimilarity between them might have been caused by the difference in the number of hours of phonetics.

Although the study does not concern the flagship interference problems and focuses on rare words and minor phonetic issues, the findings might be useful for tertiary-school teachers working with future English specialists.

There are several practical implications of this research. For example, as an immediate 'remedy' to the problems observed, more attention in the form of explicit explanation or preparation of communicative tasks could be given to the words that are still mispronounced by the majority of the group, including: 20 phonetically difficult words (*chassis*, *courteous*, *ancient*, *sergeant*, *thorough*, *disciple*, *archives*, *feud*, *prayer*, *pint*, *Niagara*, *yolk*, *developed*, *nauseous*, *risen*, *draught*, *satire*, *protein*, *suite* and *scarce*; 4 contracted forms: *don't*, *won't*, *mustn't* and *you're)*; 2 weak forms (*has* and *of*); problematic letters (<o>, i.e., representing GOAT in *folk* and STRUT in *oven*); and early lexical stress (e.g. *characterize*). The overall progress in contracted forms was the lowest (8%), therefore the teaching method concerning this category should be reconsidered.

Some other words such as *draught, mustn't, protein* and *scarce* call for improvement because non-significant progress means the course has not affected their fossilized enunciation. Thus, among the teaching methods, apart from typical listen-and-repeat exercises and transcription of individual words, learners should be exposed to high variability input by means of utilizing such free online services as *YouGlish*, *playphrase.me*, *Yarn* and *Forvo* (Appendix 3). Needless to say, there should be more spontaneous and less controlled practice of the said words in context, prepared, e.g. with the use of British National or American Cocoa corpora pages with collocates and concordance lines.

It is possible that the pronunciation of some words such as *gnaw* <*gn*> (36%) and *folk* <*l*> (32%) might have significantly changed for the better because of the participant's awareness of some spelling-governed pronunciation learnt during instruction. This means that in the classroom context we plan to continue teaching spelling-to-sound correspondence related to observed mispronunciations, such as *thorough* like Dickerson's (2015) condensed graphic rule concerning the sounds represented by <th>: 1) thV^{f} =/ð/; 2) thern/.=/ð/; 3) V/rth+E=/ð/; 4) th$_{ew}$=/θ/ (see Nowacka's (2018a) examination of spelling-to-sound correspondences in authentic materials to stimulate students' phonetic awareness).

One can speculate whether or not progress in pronunciation by 8%, as in contractions, or by 25% in words commonly mispronounced, after a two-semester course should be regarded as success. It seems that phonetic instruction left a mark on the participants' performance. The question that arises is what changes ought to be introduced in teaching these aspects to future students to help them remember what they studied and to be able to use accurate pronunciation whenever required without returning to a fossilized version.

Some implications of spelling-pronunciation research are more widely generalizable and explain how they can be applied in the teaching and research of teachers of English working with students of different L1s in other countries. The results of Nowacka's (2018b) study confirm the necessity for explicit instruction on the regularity of English spelling to eradicate pronunciation errors in the speech of 240 university students with six different L1s (Kazakh, Malaysian, Polish, Tajik, Turkish, and Ukrainian). The avoidable errors which have turned out to be the most numerous in the production task included such areas of English phonotactics as:

- the letters <-old> and <oll>,
- 'mute consonant letters' (all 6 L1s),
- two categories related to the reduction of unstressed syllables (the vowel in stress-adjacent syllables and in syllables following the stressed one to /ə/ or /ɪ/')
- 'reduced <-ous>, <-age>, and <-ate> in nouns and adjectives' (all 6 L1s)
- 'isolated errors'.

If spelling-to-sound relations are part of pronunciation training, the strain on the part of the students of memorizing phonetically challenging pronunciation exceptions will be reduced, including the ambiguous letter <o> (all 6 L1s), words with unpredictable pronunciation (all 6 L1s) and three 'unpredictable' categories: <-ough>', pronunciation of single vowel letters (all 6 L1s), and stress placement.

ACKNOWLEDGMENTS

My deepest thanks go to all 91 volunteers, the first-year students of the English Department (2017-2020) at the University of Rzeszow, Poland for participating in the research. I am also grateful to my friend Katarzyna Laskowska-Chudy for her assistance with the statistics.

ABOUT THE AUTHOR

Marta Nowacka is an Assistant Professor of Applied Linguistics at the University of Rzeszów in Poland. Her research concentrates on Polish-accented speech of university students of English and methodology of teaching phonetics to adults in EFL context. She has co-authored two course-books on English pronunciation: *How Much Wood Would a Woodchuck Chuck?*: *English pronunciation practice book* (Mańkowska et al., 2009) and *Sally Meets Harry: Primer to English Pronunciation and Spelling* (Nowacka et al., 2011).

Contact information:
Institute of English, University of Rzeszów
Kopisto 2B
35-315 Rzeszów, Poland
martha.nowacka@gmail.com

REFERENCES

Bryła-Cruz, A. (2016) *Foreign accent perception: Polish English in the British ears*. Newcastle upon Tyne: Cambridge Scholars Publishing.

Collins, B., & Mees, I. M. (2008). *Practical phonetics and phonology*. New York: Routledge.

Dickerson, W. B. (2015). Using orthography to teach pronunciation. In M. Reed & J. M. Levis (Eds.), *The handbook of English pronunciation* (pp. 488–504). Chichester: John Wiley.

iWeb: The 14 billion Word Web corpus. (n.d). Retrieved February 19, 2019, from: https://corpus.byu.edu/iweb/

Lecumberri, M. L. G., & Maidment, J. A. (2000). *English transcription course: A practical introduction*. London: Hodder Arnold.

Lewis, J. W. (n.d.). *Weakform words and contractions for the advanced EFL user*. Retrieved November 15, 2018, from http://www.yek.me.uk/wkfms.html

Mańkowska, A., Nowacka, M., & Kłoczowska, M. (2009). *How much wood would a woodchuck chuck?": English pronunciation practice book*. Rzeszów: Konsorcjum Akademickie.

Nowacka, M., Cymbalista, P., & Kleparski, G. A. (2011). *Sally meets Harry: Practical primer to English pronunciation and ppelling*. Rzeszów: Wydawnictwo Uniwersytetu Rzeszowskiego.

Nowacka, M. (2016). English spelling among the top priorities in pronunciation teaching: Polglish local versus global(ised) errors in the production and perception of *words commonly mispronounced*. *Research in Language, 14*(2), 123–148, DOI: 10.1515/rela-2016-0002.

Nowacka, M. (2018a). Phonetic inspirations in authentic materials: stimulating students' phonetic awareness. In J. Violin & R. Skarnitzel (Eds.), *The Pronunciation of English by speakers of other languages* (pp. 278–300). Newcastle upon Tyne: Cambridge Scholars Publishing.

Nowacka, M. (2018b). Back to orthoepia – spelling in pronunciation instruction: "Words commonly mispronounced" by learners of six L1s. *Research in language, 16*(4), 451–470, DOI: https://doi.org/10.2478/rela-2018-0022.

Porzuczek, A. (2015). Handling global and local English pronunciation errors. In E. Waniek-Klimczak & M. Pawlak (Eds.), *Teaching and researching the pronunciation of English: Studies in honour of W. Sobkowiak, second language learning and teaching* (pp. 169–187). London: Springer.

Roach, P. (2009). *English phonetics and phonology: A practical course*. (4th ed.). Cambridge: Cambridge University Press.

Sobkowiak, W., & Szpyra-Kozłowska, J. (2001). *Workbook in English phonetics*: *for Polish students of English*. Lublin: Wydawnictwo UMCS.

Sobkowiak, W. (1996). *English phonetics for Poles*. Poznań: Bene Nati.

Szpyra-Kozłowska, J. (2013). On the irrelevance of sounds and prosody in foreign-accented speech. In E. Waniek-Klimczak & L. Shockey (Eds.), *Teaching and researching English accents in native and non-native speakers* (pp. 15–29). Berlin: Springer.

Szpyra-Kozłowska, J. (2015). *Pronunciation in EFL instruction: A research-based Approach*. Bristol: Multilingual Matters.

Upward, C. (1994). The classic concordance of cacographic chaos. *Journal of the Simplified Spelling Society, 17*(2), 27–31.

Wells, J. C. (1982). *Accents of English I*: *An introduction*. Cambridge: Cambridge University Press.

Wells, J. C. (2008). Goals in Teaching English Pronunciation. In K. Dziubalska-Kołaczyk & J. Przedlacka (Eds.), *English pronunciation models: A changing scene*. (2nd ed.). (pp. 101–110). Bern: International Academic Publishers.

Zając, M. (2015). Compiling a corpus-based list of words commonly mispronounced. In E. Waniek-Klimczak & M. Pawlak (Eds.), *Teaching and researching the pronunciation of English* (pp. 153–167). Heidelberg: Springer.

Appendix 1. Diagnostic Test.

Task 1: word reading

1. saw	2. sweat	3. thorough	4. abroad	5. ghoul	6. cough	7. word	8. suite
9. threat	10. hawk	11. ancient	12. folk	13. satire	14. courteous	15. sergeant	16. protein
17. sponge	18. scarce	19. failure	20. author	21. accurate	22. comb	23. purchase	24. area
25. foreign	26. oven	27. steak	28. fiend	29. disciple	30. gnaw	31. archives	32. chassis
33. draught	34. feud	35. charlatan					

Task 2: sentence-reading

1. The sun has just risen.	11. It won't make sense.
2. She said that she would come.	12. Aren't you appalled?
3. Yolk isn't a white part of an egg.	13. You mustn't lay it on the floor.
4. Some people say English pronunciation is difficult.	14. He's the most determined player I have ever seen.
5. I would like to see Niagara Falls one day.	15 The basic course is not available.
6. My dad prepares the best pizza dough.	16. We need to develop a European rail network.
7. Don't draw a leopard on these walls.	17. I feel nauseous.
8. I swear I must have lost it.	18. What is chaos?
9. A pint of lager please.	19. Oh God hear my prayer.
10. You're the apple of my eye.	20. That is the end of the test, thank you.

Appendix 2. Words and phonetic aspects examined in Task 1 and 2 (source, error, accepted pronunciation).

Table 1.

Task 1

<table>
<tr><th>No.</th><th>Word[iv]</th><th>Source</th><th>Examined aspect[v]</th><th>Example of an error in this study</th><th>Example of accepted pronunciation</th></tr>
<tr><td>1.</td><td>abroad</td><td>WCM</td><td><oa> ≡ THOUGHT || PALM</td><td>/əˈbrəʊt, e-/</td><td>/əˈbrɔːd/ || /əˈbrɑːd/[vi]</td></tr>
<tr><td>2.</td><td>accurate</td><td>WCM</td><td><-ate> as /ət/ and stress on the 1st syllable (henceforth syll.).</td><td>/ˈækjʊreɪt/</td><td>/ˈækjərət/, /-jʊr-/, /-t/</td></tr>
<tr><td>3.</td><td>ancient</td><td>WCM</td><td>whole word (FACE followed by /n/)</td><td>/ˈenʃɪnt/, /ˈeɪʃɪnt/</td><td>/ˈeɪnʃənt/</td></tr>
<tr><td rowspan="2">4.</td><td>archives_W</td><td>spelling</td><td>whole word (START, /k/ PRICE)</td><td>/ɑː(r)ˈtʃiːfs/</td><td rowspan="2">/ˈɑː(r)kaɪvz/</td></tr>
<tr><td>archives<ch></td><td>spelling</td><td><ch>≡/k/</td><td><ch>=/tʃ/,<ch>=/ʃ/ /ɑː(r)ˈ(t)ʃiːfs/</td></tr>
<tr><td>5.</td><td>area</td><td>WCM</td><td>whole word (stress on the 1st syll., SQUARE || DRESS/TRAP)</td><td>/əˈriə/</td><td>/ˈeəriə/ || /ˈeriə/, /ˈæriə/</td></tr>
<tr><td>6.</td><td>author</td><td>WCM</td><td><au>≡THOUGHT || PALM</td><td>/ˈaʊθə, ˈəʊ-/</td><td>/ˈɔːθə/ || /ˈɔːθər/, /ɑː- /</td></tr>
<tr><td rowspan="2">7.</td><td>charlatan_COMMA</td><td>spelling</td><td>COMMA in the 3rd syll.</td><td>/ˈʃɑːlʌtʌn/</td><td>/ˈʃɑːlətən, -æn / || /ˈʃɑːrlətᵊn/</td></tr>
<tr><td>charlatan<ch></td><td>spelling</td><td><ch>≡/ʃ/</td><td>/ˈtʃɑːrlətᵊn/</td><td>/ˈʃɑːlətən, -æn/ || /ˈʃɑːrlətᵊn/</td></tr>
<tr><td rowspan="2">8.</td><td>chassis_W</td><td>spelling</td><td>whole word (stress on the 1st syll., <ch>≡ /(t)ʃ/, TRAP, /-si/)</td><td>/ˈtʃeɪsɪs, ˈtʃæ-/, /tʃæˈziːs/</td><td>/ˈʃæsi(z)/ || /ˈtʃæsi(z)/</td></tr>
<tr><td>chassis<ch></td><td>spelling</td><td><ch>≡/ʃ/, /tʃ/</td><td>/ˈkɑːsɪs/</td><td>/ˈʃæsi/ || /ˈtʃæsi/</td></tr>
<tr><td>9.</td><td>comb</td><td>WCM</td><td>silent <b></td><td>/kɒmp, kʌ-/</td><td>/ˈkəʊm/ || /ˈkoʊm/</td></tr>
<tr><td>10.</td><td>cough</td><td>WCM</td><td><ough>≡LOT, THOUGHT || PALM and /f/</td><td>/kɒt/, /kɔːt/</td><td>/ kɒf/, /kɔːf/ || /kɔːf/, /kɑːf/</td></tr>
</table>

11.	*courteous*	WCM	whole word (NURSE, NORTH followed by /tiəs/)	/ˈkɔ:(r)tʃəs, ˈkɜ:-/	/ˈkɜ:tiəs/, /ˈkɔ:tiəs/ || /ˈkɜ:rt̬iəs/
12.	*disciple*	Chaos	whole word (stress on the 2nd syll., PRICE, weak syllables	/ˈdɪsɪp(ɪ)l, (-ə)l/	/dɪˈsaɪpᵊl, də-/
13.	*draught*	challenging word	whole word (BATH followed by /ft/)	/drɔ:t/	/drɑ:ft/, /-æ-/ || /dræft/
14.	*failure*	WCM	whole word (FACE, /j/ in a 2nd syll.)	/ˈfeɪlə(r)/	/ˈfeɪljə/ || /ˈfeɪljər/
15.	*feud*	Chaos	whole word (GOOSE preceded by /j/)	/fəʊd/	/fju:d/
16.	*fiend*	Chaos	FLEECE	/faɪnd/	/fi:nd/
17.	*folk_V*	WCM	GOAT	/fɒlk/	/fəʊk/, /fəʊlk/[vii]
	folk<l>	WCM	silent <l>		
18.	*foreign_S*	WCM	stress on the 1st syll.	/fɒˈreɪn/	/ˈfɒrən, -ɪn/ || /ˈfɔ:rən, ˈfɑ:-/
	foreign_W	WCM	LOT ||NORTH/START, SCHWA/KIT		
19.	*ghoul*	challenging word	GOOSE	/gɔ:l/	/gu:l/
20.	*gnaw_V*	spelling	THOUGHT ||PALM	/nəʊ/, /naʊ/	/nɔ:/ || /nɑ:/
	gnaw<gn>	spelling	silent <gn>	/gnɔ:/	
21.	*hawk*	WCM	THOUGHT || PALM	/həʊk/	/hɔ:k/ || /hɑ:k/
22.	*oven*	WCM	<o>≡STRUT	/ˈəʊən/, /ˈəʊv(ə)n/, /ˈɒv(ə)n/, /ˈaʊən/	/ˈʌvᵊn/
23.	*protein_S*	WCM	stress on the 1st syll.	/prəʊˈti:n/, /prəˈteɪn/, /prɒˈti:n/	/ˈprəʊti:n/, /ˈprəʊti:ɪn/ || /ˈproʊ-/
	protein_V	WCM	<o>≡GOAT, <ei>≡FLEECE	/prəˈteɪn/, /prɒˈti:n/, /ˈprɒteɪn/	
24.	*purchase_S*	WCM	stress on the 1st syll.	/pə(r)ˈtʃeɪs/	
	purchase_W	WCM	whole word (NURSE, SCHWA/KIT)		

25.	*satire*	WCM	whole word (stress on the 1st syll., TRAP, triphthong: PRICE + SCHWA)	/səˈtaɪə/	/ˈsætaɪə/ ‖ /ˈsætaɪər/
26.	*saw*	WCM	THOUGHT ‖ PALM	/səʊ/	/sɔː/ ‖ /sɑː/
27.	*scarce*	WCM	SQUARE ‖ DRESS/TRAP	/skɑːrs/	/skeəs/ ‖ /ske^{ə}rs/, / skæ^{ə}rs/
28.	*sergeant*	WCM	whole word (START, <ge>≡/dʒ/)	/ˈsɜːrdʒənt/	/ˈsɑːdʒənt/ ‖ /ˈsɑːrdʒənt/
29.	*sponge*	WCM	STRUT	/spɒntʃ/	/spʌndʒ/
30.	*steak*	WCM	FACE	/stek/, /stiːk/, /stɪk/	/steɪk/
31.	*suite*	challenging word	FLEECE preceded by /sw/	/s(j)uːt/[viii]	/swiːt/
32.	*sweat*	WCM	DRESS	/swiːt/	/swet/
33.	*thorough*	WCM	whole word (STRUT, SCHWA ‖ NURSE, GOAT)	/ˈθɔːrəʊ/	/ˈθʌrə/[ix] ‖ /ˈθɜːroʊ/[x]
34.	*threat*	WCM	DRESS	/θriːt/, /θrɪt/	/θret/
35.	*word*	WCM	NURSE	/wɔːrt/	/wɜːd/ ‖ /wɜːrd/

Table 2. Task 2

<table>
<tr><th>No.</th><th>Word</th><th>Source</th><th>Examined aspect</th><th>Example of an error</th><th>Example of accepted pronunciation</th></tr>
<tr><td>1.</td><td>don't</td><td rowspan="4">contracted form list</td><td>GOAT</td><td>/dɒnt/</td><td>/dəʊnt/</td></tr>
<tr><td>2.</td><td>mustn't</td><td>whole word (STRUT, /s/)</td><td>/mʌznt/</td><td>/ˈmʌsᵊnt/</td></tr>
<tr><td>3.</td><td>won't</td><td>GOAT</td><td>/wɒnt/</td><td>/wəʊnt/</td></tr>
<tr><td>4.</td><td>you're</td><td>NORTH /CURE</td><td>/ˌju ˈɑ:(r)/</td><td>/jɔ:r, jʊər/</td></tr>
<tr><td>5.</td><td>has</td><td rowspan="7">weak/strong form list</td><td rowspan="4">unstressed form, COMMA</td><td>/hæz, -s/</td><td>/(h)əz/</td></tr>
<tr><td>6.</td><td>have</td><td>/hæv/</td><td>/(h)əv/</td></tr>
<tr><td>7.</td><td>of</td><td>/ɒv, -f/</td><td>/əv/</td></tr>
<tr><td>8.</td><td>that</td><td>/ðæt/</td><td>/ðət/</td></tr>
<tr><td>9.</td><td>the</td><td>/ði/ before a vowel</td><td>/ðə/</td><td>/ði/</td></tr>
<tr><td>10.</td><td>would</td><td>unstressed form with COMMA</td><td>/wʊd/</td><td>/(wə)d/</td></tr>
<tr><td>11.</td><td>some</td><td>strong form</td><td>/sɒm/</td><td>/sʌm/</td></tr>
<tr><td>12.</td><td>appalled</td><td>WCM modified</td><td>whole word (stress, THOUGHT)</td><td>/ʌˈpɑ:ld, e-, ə-/[xi], /eˈpi:lt, ə-, a-, ˈpɪ-/</td><td>/əˈpɔ:ld ‖ əˈpɑ:-/</td></tr>
<tr><td>13.</td><td>available</td><td>WCM</td><td>whole word (stress, FACE, unstressed syllables)</td><td>/ʌˈvaɪəb(ɪ)l, e-ˈ, eˈveɪ-/, /eˈveləb(ɪ)l, -li-/</td><td>/əˈveɪləbᵊl/</td></tr>
<tr><td>14.</td><td>basic</td><td>WCM</td><td>whole word (FACE, /s/)</td><td>/ˈbeɪzɪk/</td><td>/ˈbeɪsɪk/</td></tr>
<tr><td>15.</td><td>determined</td><td>WCM</td><td>whole word (stress, NURSE)</td><td>/ˈdetəmaɪnd/, /dəˈtɜ:(r)mɪneɪtɪd/, /dəˌtɜ:(r)mɪnˈeɪtɪd/</td><td>/dɪˈtɜ:(r)mɪnd, də-, -ənd/</td></tr>
<tr><td>16.</td><td>nauseous</td><td>Chaos</td><td>whole word (stress, THOUGHT | | PALM, unstressed syllables)</td><td>/ˈnɔ:sɪs, -ʊs ˈnɔ:zəs, -s-/ /ˈnɔɪz(j)əs, -es/, /ˈn(j)u:ʃəs/ /ˈnu:ʃɒs/ - extremely varied</td><td>/ˈnɔ:sɪəs, -z- | | ˈnɔ:ʃəs, ˈnɑ:ʃ-; ˈnɔ:zɪəs, ˈnɑ:z-/</td></tr>
<tr><td>17.</td><td>chaos</td><td>Chaos</td><td>whole word (<ch>≡/k/, FACE, LOT | | PALM</td><td>/ˈkɑ:ɒs, ˈh-/, /ˈkeɒs/</td><td>/ˈkeɪɒs | | -ɑ:s/</td></tr>
<tr><td>18.</td><td>course</td><td>‘Woodchuck’ coursebook</td><td>NORTH</td><td>/kɜ:(r)s/</td><td>/kɔ:(r)s/</td></tr>
</table>

19.	*dough*	WCM	GOAT	/daʊ/, dɔ:/	/dəʊ/
20.	*lager*	‘Woodchuck’ coursebook	whole word (PALM, <g> ≡ /g/)	/ˈlɑ:dʒə/, /ˈleɪdʒə/	/ˈlɑ:gə/
21.	*leopard*	WCM	whole word (stress, DRESS, unstressed syll.)	/ˈlɪəpɑ:d/	/ˈlepə(r)d/
22.	*Niagara*	‘Woodchuck’ coursebook	whole word (stress, TRAP, unstressed syll.)	/nɪəˈgɑ:rə/, /ˈniɑ:gərə/	/naɪˈæg(ə)rə, ni-/
23.	*pint*	WCM	PRICE	/pɪnt/	/paɪnt/
24.	*prayer*	Chaos	SQUARE ‖ DRESS / TRAP	/ˈpreɪə/	/preə | | preᵊr, præᵊr
25.	*pronunciation*	Chaos	whole word (stress, STRUT, unstressed syll.)	/prəˌnaʊsiˈeɪʃən/	/prəˌnʌnsiˈeɪʃən/
26.	*yolk*	WCM	GOAT	/jɒlk/	/jəʊk/[xii]
27.	*develop*	WCM	whole word (stress, DRESS, unstressed syll.)	/ˈdevəlɒp, di:-, dɪ-, - vɪ-/ , /ˈdevələʊp/ /ˈdevələʊp/	/dɪˈveləp, də-/
28.	*draw*	irregular verb list/WCM	THOUGHT | | PALM	/drəʊ/, /draʊ/	/drɔ: | | drɑ:/
29.	*lay*	irregular verb list	FACE	/laɪ/	/leɪ/
30.	*risen*	irregular verb list	KIT	/ˈraɪzən/	/ˈrɪzᵊn/

Appendix 3

Task A. English versus Polglish (*YouGlish*, *playphrase.me*, *Yarn* and *Forvo*)

Record your pronunciation of the following words:
ancient, *archives*, *characterize*, *chassis*, *courteous*, *developed*, *disciple*, *don't*, *draught*, *feud*, *folk*, *has*, *mustn't*, *nauseous*, *Niagara*, *of*, *oven*, *pint*, *prayer*, *protein*, *risen*, *satire*, *scarce*, *sergeant*, *suite*, *that*, *this*, *thorough*, *won't*, *yolk* and *you're*.

Then listen to them on: *YouGlish*, *playphrase.me*, *Yarn* or *Forvo*. Repeat the phrases after speakers. Compare your own pronunciation of these words with the one by native speakers. Transcribe the above-mentioned words.

Questions: Does your pronunciation of these words agree with the one you heard? If not, in what way does is differ? Is your pronunciation an example of mispronunciation or a variant form used in one variety of English?

Write down your answers.

---

[i] Wells' (1982) standard lexical sets for vowels are applied.

[ii] Rank order is marked with #. It shows which place a specific word occupies in the corpus, e.g. the rank order of #153 for *area*, means that it is the 153rd most frequent word of the 60,000 most frequent words in this corpus.

[iii] iWeb's note on word frequency: "high frequency words (about word #5000 in the 60,000-word list), medium frequency (~25,000), and low frequency (~45,000) words.
On the basis of the above word frequency ranges the following scale regarding word frequency was applied in the present study: 1 – 14,999 high frequency words; 15,000 – 34,999 medium frequency words; 35,000 – 60,000 low frequency words.

[iv] The letters and symbols which are used next to some words stand for: < > a letter included between triangular brackets concerns the rendition of this letter or letters, e.g. gnaw <gn> regards the silent letter <g> in *gnaw* /nɔː/ or <l> in *folk*; _S: lexical stress, _V: a stressed vowel, e.g. THOUGHT in *gnaw*,_W: the pronunciation of a whole word, stressed and unstressed vowels, e.g. *purchase* /ˈpɜːtʃəs/.

[v] The meaning of symbols used in this section is as follows: <>≡ a spelling correspondence between a letter/letters and a phoneme/phonemes, || the difference in pronunciation between British (on the left of the double lines) and American English (on the right side).

[vi] Wells (2000) notes that *abroad* exhibits pronunciation unexpected for this spelling.

[vii] The non-standard pronunciation with /l/ was not accepted.

[viii] The pronunciation /suːt/ was regarded as a mistake although Wells (2000) reports that in AmE *suite* can be pronounced as /suːt/ in the sense '*suite of furniture*'.

[ix] The initial consonant 'eth' was not the focus of this assessment. Students were not penalised for substituting theta with /f, t/ therefeore /ˈfʌrə/ (n=8) and /ˈtʌrə/ (n=1) was accepted as correct.

[x] Wells (2000) makes a comment that in *thorough* RP and GenAm differ in an unpredictable and striking way.

[xi] The American pronunciation of a word *appalled* is /əˈpɑːld/. The analysis was controlled for the British THOUGHT. If the enunciations /aˈpalt, e-, ə-/[xi] (26%) have been accepted, it would give rise to the higher overall score of 66% for the word *appalled*.

[xii] The silent /l/ in *yolk*, which is characteristic for British English, was included. However, this sound in the word is not a mistake in some American varieties /joʊk, joʊlk, jelk/.

Ruellot, V. (2019). French stereotypical accent and pronunciation development of /p/, /t/, and /k/. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Ames, IA, September 2018 (pp. 341-353). Ames, IA: Iowa State University.

PRESENTATION/POSTER

FRENCH STEREOTYPICAL ACCENT AND PRONUNCIATION DEVELOPMENT OF /P/, /T/, AND /K/

Viviane Ruellot, Western Michigan University

This study aims to contribute to research on second language (L2) accent imitation in the native language and its relation to L2 pronunciation development (e.g., Everitt, 2015; Rojczyk, Porzuczek, & Bergier, 2013). Unlike most previous studies, it focuses on stereotypical—rather than authentic—accent and its potential benefits (salience and learner familiarity) for pronunciation improvement. Over three weeks, 14 American learners of French in three groups received pronunciation instruction and based their practice on models speaking English either with a stereotypical French accent (n=5) or an authentic French accent (n=4), or on models who were native speakers of French speaking French (n=6). Learners and native speaker controls recorded their pronunciation of texts they read before and after practice. Words featuring /p/, /t/, or /k/ in initial position were selected and voice onset time (VOT) of the plosives was measured. A subset of the same words was presented to native speaker raters for assessment of accentedness. While listener ratings yield differences that are significant or approaching significance between the control group and the experimental groups, VOT measures do not do so. Results are discussed in terms of perception of accentedness versus acoustic measurement.

INTRODUCTION

Second language (L2) pronunciation has traditionally been taught and learned through the intuitive-imitative approach (Celce-Murcia, Brinton, & Goodwin, 1996). Learners listen to models speaking the L2 and repeat after them, trying to imitate the modeled pronunciation to the best of their ability. Recently, research interest in pronunciation imitation has resurfaced… with a twist: Instead of imitating the L2 accent in the L2, subjects have been asked to imitate the L2 accent while speaking in their native language (L1). This approach was used in several studies, either as a way to measure learner awareness of certain L2 pronunciation features (Flege & Hammond, 1982; Mora, Rochdi, & Kivistö-de Souza, 2014; Neuhauser, 2011; Rojczyk, 2012, 2015; Rojczyk et al., 2013; Sypiańska & Olender, 2013), or as a tool to learn L2 pronunciation (Everitt, 2015). The latter approach is also used in the present study. However, it itself puts a twist on the notion of L2 accent imitation in the L1 by using, as the model for practice, a stereotypical version of the L2 accent. The exaggeration of pronunciation features in a stereotypical accent, as well as the unconscious knowledge about stereotypical accents learners have accumulated through exposure to the media, may make the use of L2 stereotypical accent imitation in the L1 a valuable tool for the teaching and learning of L2 pronunciation. This study is part of a larger one investigating the impact of this approach on the development of several features of French pronunciation. The results for French /ʁ/ were reported in Ruellot (2018). The present study focuses on the French voiceless stops /p/, /t/, and /k/ and tests the impact of training with stereotypical accent on the reduction of their aspiration.

BACKGROUND

Imitation is fundamental to human learning (Nagell, Olguin, & Tomasello, 1993) and remains so all along the lifetime. From learning to tie one's shoes with the guidance of an adult, to fixing a furnace following one of the many tutorial videos available online, the acquisition of skills involves some form of mimicry. Similarly, imitation has always been central to L2 pronunciation acquisition, even when analytic elements, such as the study of phonetics, were incorporated in the late 19th century as a supplement to imitation (Kelly, 1969). With this seemingly simple approach, learners listen to a model speaking the L2 and then repeat what they heard in their best pronunciation. Not only must learners draw on their perception and production skills to replicate the modeled pronunciation, but they also must process—at least to some extent—the grammar and vocabulary in the speech sample to make sense of what they are repeating. The latter aspect of replication mobilizes a part—large or small, depending on the learners' proficiency level—of their working memory, and reduces the resources therein that could be otherwise dedicated to the processing of pronunciation information.

An approach that allows to maximize allocation of processing resources is one where the content to be imitated is not in the target language but in the learner's native language. The imitation of a foreign accent in the native language has been the focus of recent studies in the acquisition of L2 pronunciation. It was used to measure learner awareness of L2 stop consonant voice onset time (VOT), i.e., the length of time between the release of a plosive and the onset of voicing (Lisker & Abramson, 1964). This approach proved effective in assessing learner awareness—both implicit and conscious—that the VOT of /p/, /t/, and /k/ is longer in L2 English than it is in L1 Spanish (Mora et al., 2014) and in L1 Polish (Sypiańska & Olender, 2016), and that it is shorter in L2 Spanish than in L1 English (Flege & Hammond, 1982) and in L2 French compared to L1 German (Neuhauser, 2011). It was also used as a learning tool to improve L2 pronunciation, with mixed results. The approach did not prove effective in helping L1 Anglophone learners of L2 French improve their pronunciation of French /ʁ/ in Ruellot (2018). But it effectively helped L1 Spanish learners increase aspiration of English /p, t, k/ in Everitt (2015). This study set out to determine the extent to which stereotypical accent imitation could also help American learners decrease aspiration of French /p, t, k/. While excessive aspiration may not significantly affect the comprehensibility and intelligibility of L2 French speakers, it remains a strong marker of a foreign accent (Major, 1987; Flege & Eefting, 1987; Riney & Takagi, 1999). Reducing the aspiration of /p, t, k/, by keeping muscles tense to limit air release, might prove easier than improving pronunciation of French /ʁ/, which involves a configuration of articulators (i.e., drawing back the tongue to form a pharyngeal, velar, or uvular constriction) that is absent from the English repertoire.

Using stereotypes for teaching purposes might not *a priori* be considered favorably by pedagogues. After all, stereotypes offer an incomplete and reductive perspective of a people and its culture by selecting some aspects of that culture and grossly exaggerating them. This is also true of stereotypical accents. However, that exaggeration, which renders pronunciation features more salient (Kristiansen, 2003), may in fact help learners become aware of those features so they can begin to process and acquire them (Schmidt, 1990, 1993). Furthermore, learners are generally familiar with stereotypical accents and their characteristics, as they were exposed to them from a young age through animated films and other media (Lippi-Green, 1997) featuring so-called French characters, such as Warner Brothers' Pepé Le Pew and Steve Martin's Inspecteur Cluzot in *The Pink Panther* (Simonds & Levy, 2006). Harnessing this unconscious knowledge may benefit L2 pronunciation learning of certain features, including the French voiceless stop consonants.

As mentioned above, awareness and development of aspiration of voiceless plosives have mostly been assessed with acoustic measures of VOT (Everitt, 2015; Flege & Hammond, 1982; Mora et al., 2014; Neuhauser, 2011; Rojczyk, 2012; Rojczyk et al., 2013; Sypiańska & Olender, 2013). However, some scholars in the field of pronunciation caution against limiting assessment to acoustic measurement. Thomson and Derwing (2014: 337) explain that "measurable changes are not always noticeable to listeners." Indeed, a study investigating pronunciation acquisition in a child presenting phonological disorders (Maxwell & Weismer, 1982) showed that while the child produced statistically different VOTs for voiced and voiceless stops, the judges perceived all the stops as voiced. As Thomson and Derwing (2014) advise that "in the final analysis, it is what listeners perceive that matters" (p. 337), both VOT measures and native speaker judgements were included in the present study, which was guided by the following research questions:

1. Does practice speaking L1 English with a French stereotypical accent help significantly reduce aspiration of voiceless plosives in L2 French production?
2. Do VOT measures and French native speaker judgments correlate?

METHOD

Participants

Fourteen intermediate students, enrolled in a French pronunciation course at a university in the US, practiced French pronunciation in one of three groups: the Stereotypical accent group (n=5), modeled by native speakers of English speaking English with a French stereotypical accent; the Authentic accent group (n=4), in which the models were French natives speaking English with an authentic (i.e. not exaggerated) French accent; and the French accent group (n=6), which followed the traditional approach of basing pronunciation study and practice on French native speakers speaking French. Six French native speakers additionally participated in the study as controls.

Treatment & sounds

Students took part in three in-class twenty-minute training sessions during which they received explicit instruction (e.g., articulation information) on the following French features: /ʁ/, the front vowels [ø] and [y], vowel stability (i.e., lack of reduction), intonation, and reduced aspiration of /p, t, k/. Although stereotypical accents are familiar to people, pronunciation characteristics remain general and imprecise to most listeners (Honey, 2017). This is why explicit instruction was included in the design of this study, and for all groups so as not to put any one at an advantage. After each session, students practiced their pronunciation of the five features at home by recording themselves imitating their assigned model speaking five sentences. They repeated each sentence at least three times before recording themselves.

Tests and data assessment

All groups were tested before, immediately after, and one week after treatment, reading the same narrative and dialogue in French. For the present study, fourteen words with initial /p, t, k/ were extracted from the narrative and the dialogue at all three times (Table 1). Word initial /p, t, k/ were chosen because of the greater impact of the mispronunciation of sounds occurring at the beginning of a word (Flege & Munro, 1994). The consonants' VOT was measured in Praat (Boersma & Weenink, 2018) by the researcher. Because VOT varies as a function of speech rate, a ratio of consonant to syllable duration was used to "normalize" the data (Summerfield, 1981). The ratio was obtained from dividing the VOT by the duration of the syllable in which the sound appeared (Boucher, 2002). Some of the words were also presented to three French native speakers who all rated the pronunciation of the initial consonant in all the words of that subset using a nine-point Likert-type accentedness scale (Derwing, Rossiter, Munro, & Thomson, 2004; Tanner & Landon, 2009) ranging from 1- Very strong foreign accent to 9- No foreign accent. Data collection is on-going and native speaker rating (NS rating) data is currently limited to 30% of the data, i.e., two of the five words with /p/ and to two of the six words with /t/ (see Table 1 below).

Table 1

Words included in the corpus

Sound	VOT Ratio	NS Rating
/p/	*par* (by), *parce que* (because), *Paris*, *passer* (pass), *pour* (for)	*Paris*, *passer*
/t/	*taches* (spots), *table*, *tapis* (rug), *temps* (time), *tous* (everyone), *tout* (everything)	*taches*, *tout*
/k/	*canapé* (sofa), *courir* (run), *quand* (when)	n/a

RESULTS

Interrater reliability and group differences before treatment

The degree of agreement between native speaker judges was calculated and found to be high: the average measures intraclass correlation coefficient was .830 with a 95% confidence interval from .799 to .857 ($F(404,808) = 5.979$, $p < .001$). A one-way ANOVA revealed no significant differences between the three groups on the pre-test, whether in their rated production or in the VOT ratios, indicating that the groups were similar before treatment (NS ratings: $F(2,11) = .659$, $p = .536$; VOT ratios: $F(2,11) = 1.297$, $p = .312$).

Impact of treatment

To assess the impact of the treatment, two repeated measures ANOVAs were run: one with the participants' VOT ratios for /p/, /t/, and /k/, and one with the NS ratings for /p/ and /t/. Both tests had *Group* as a between-subjects factor and *Time* as a within-subjects factor. Results for VOT ratios indicate no significant difference between groups ($F(3, 16) = 1.789$, $p = .190$), while NS ratings results do ($F(3, 10) = 5.727$, $p = .015$). Bonferroni-adjusted multiple comparisons (Table 2) identify the significant differences: between the native speaker controls and the Authentic and the Stereotypical groups, i.e., participants who practiced pronunciation following a model that spoke L1 English.

Table 2

Bonferroni-adjusted multiple comparisons for group (NS ratings)

Groups	Mean Diff.	SD	Sig.	95% CI for Diff.	
				Lower	Upper
Authentic - French	-1.253	.774	.818	-3.789	1.282
Authentic - Stereotypical	-.451	.774	1.000	-2.987	2.084
Authentic - Native Speakers	-3.750*	.948	.016	-6.856	-.644
French - Stereotypical	.802	.774	1.000	-1.734	3.338
French - Native Speakers	-2.497	.948	.150	-5.602	.609
Stereotypical - French	-.802	.774	1.000	-3.338	1.734
Stereotypical - Native Speakers	-3.299*	.948	.036	-6.404	-.193

* Significant at the $p < .05$ level

Results in Table 3 show a significant effect of *Time* for VOT ratios.

Table 3

Tests of within-subjects effects from the repeated-measures ANOVA on VOT ratios and on NS ratings

	VOT Ratios						NS Ratings					
Source	Type II Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Type II Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Time	0.016	2	0.008	3.803*	0.033	0.192	0.943	2	0.471	0.622	0.547	0.059
Time x Group	0.020	6	0.003	1.616	0.175	0.233	3.175	6	0.529	0.698	0.654	0.173

* Significant at the $p < .05$ level

Bonferroni-adjusted pairwise comparisons (Table 4) indicate that VOT ratios had significantly increased and were higher at delayed post-test (Post 2) than before treatment (Pre-test), suggesting negative long-term impact of treatment. However, the absence of a significant effect of *Time* for NS ratings suggests that time has no impact on the perceived quality of the pronunciation of /p/ and /t/. These results are discussed below.

Table 4

Bonferroni-adjusted multiple comparisons for tests (VOT ratios)

Test	Mean Diff.	SD	Sig	95% CI for Diff.	
				Lower	Upper
Pre - post 1	-0.011	0.008	0.532	-0.032	0.010
Pre - post 2	-.025*	0.008	0.019	-0.046	-0.004
Post 1 - post 2	-0.014	0.009	0.482	-0.039	0.011

* Significant at the p < .05 level

VOT ratios vs rater judgments

To assess the relationship between VOT ratios and native speaker ratings, a series of Pearson product-moment correlation coefficient were computed. Data collection is on-going, and the results presented here correspond to 30% of the data, which include two of the five /p/ words (*Paris*, *passer* 'to pass'), and two of the six words with /t/ (*taches* 'spots' and *tout* 'all').

Results indicate a significant negative correlation between VOT ratios and listener ratings of /p/: $r(50) = -.76$, $p < .001$, $R^2 = .58$, CI [-.86, -.62]. However, the correlation for /t/, which is also negative, is non-significant: $r(47) = -.15$, $R^2 = .02$, CI [-.42, .12]. These correlations are illustrated in Figures 3 and 4.

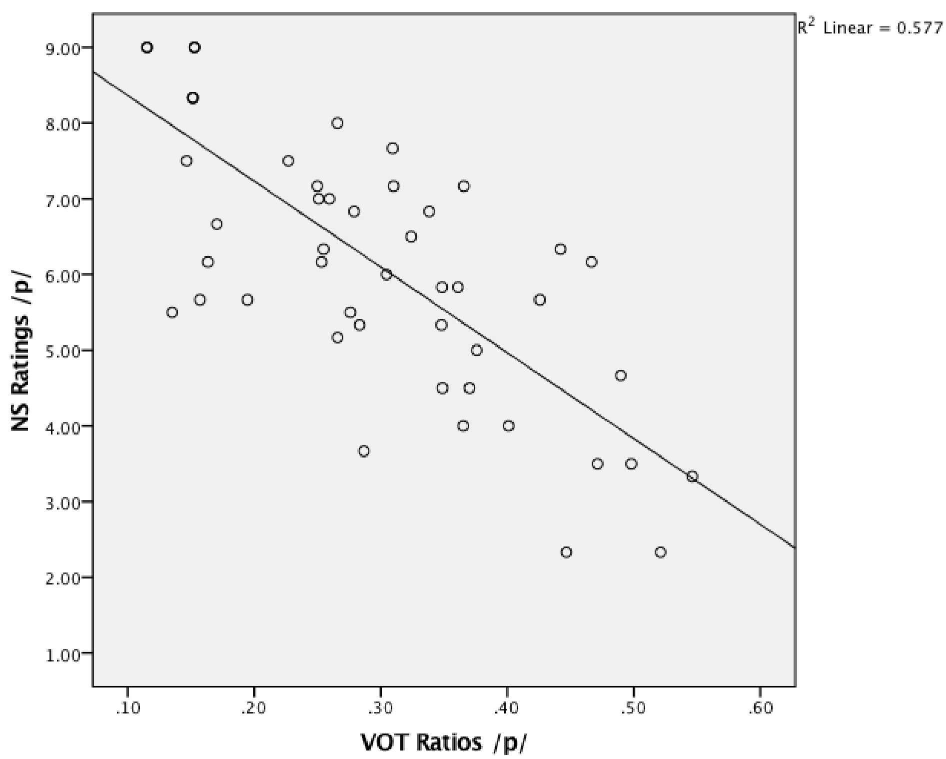


Figure 3. Scatter plot of VOT ratios vs NS ratings for /p/.

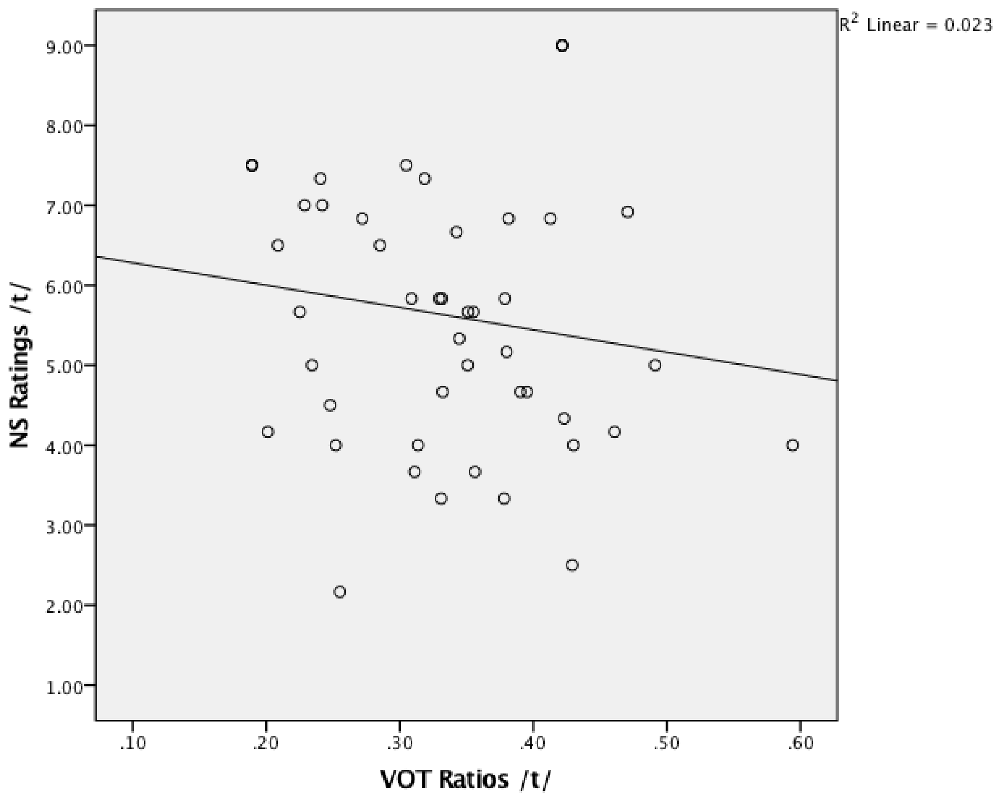


Figure 4. Scatter plot of VOT ratios vs NS ratings for /t/.

Considering that the two words in /p/ (*Paris* and *passer*) had the same environment (i.e., both /p/ followed by /a/) but not the two words in /t/ (*taches* and *tout*, i.e., /t/ + /a/ and /t/ + /u/), separate correlations were run for *taches* and *tout*. A significant correlation for *taches* productions (Figure 5) was found ($r(47) = -.78$, $p < .001$, $R^2 = .60$, CI [-.86, -.68]), but not for the *tout* productions ($r(47) = .16$, $R^2 = .02$, CI [-.18, .48] – Figure 6). The researcher listened to the participants' *tout* productions and heard friction, probably resulting from relaxed muscles and reduced vowel tension, and which may have been made even more noticeable by a vocal tract lengthened in anticipation of the /u/. This finding is further discussed below.

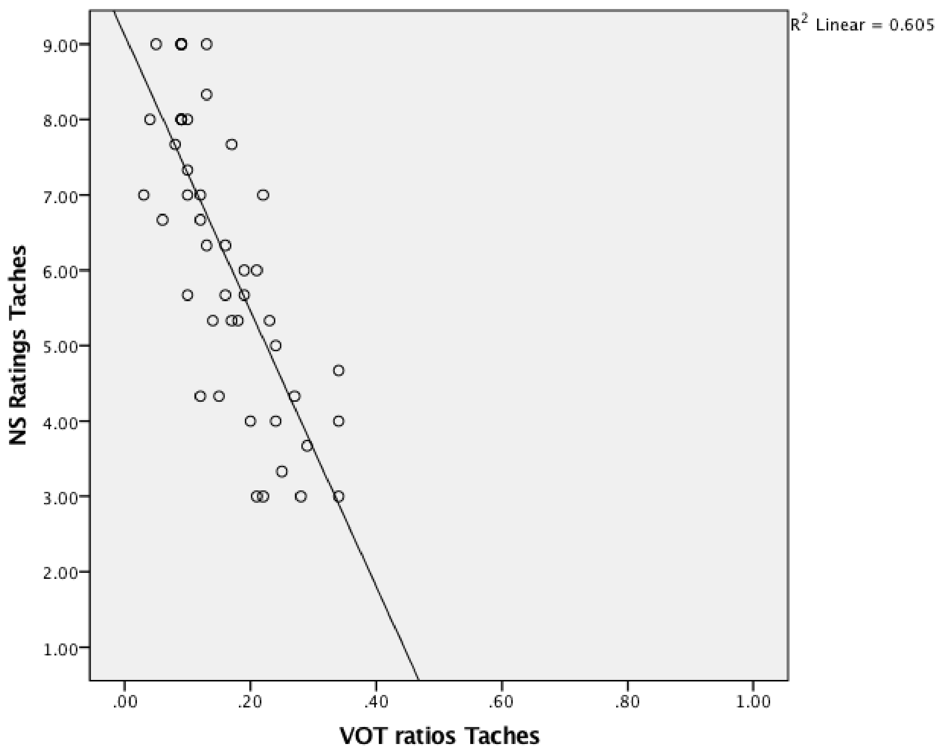


Figure 5. Scatter plot of VOT ratios vs NS ratings for taches productions.

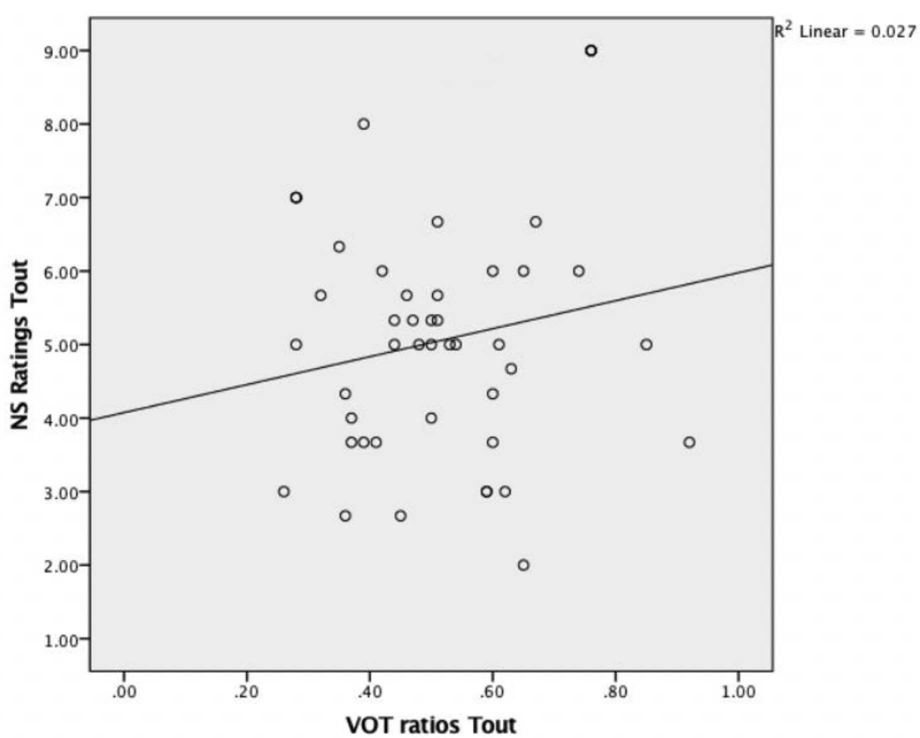


Figure 6. Scatter plot of VOT ratios vs NS ratings for tout productions.

DISCUSSION AND CONCLUSION

Tests were run to determine whether practicing speaking L1 English with a French stereotypical accent helps significantly reduce aspiration of French voiceless plosives. Both the absence of a significant difference between the experimental groups and the control group and a lack of a significant interaction between group and time for VOT ratios might suggest that learners pronounced word initial /p/, /t/, and /k/ with a native-like degree of aspiration before, immediately, and one week after treatment. In other words, as the VOT ratios of their voiceless plosives were comparable to those of the native speaker controls even before treatment, learners did not need to improve their pronunciation of those sounds. Native speaker raters, on the other hand, did not seem to agree and findings would indicate that they perceived the production of the subjects in the Authentic and the Stereotypical groups to be far from native-like before treatment, but also after. Based on native speaker judgements then, imitating an L2 accent—be it exaggerated or not—in the L1 did not help improve pronunciation of French /p, t, k/. Care needs to be taken when interpreting the latter results however, as they may be influenced by the small number of rated tokens analyzed (only four rated tokens per subject against 14 tokens with VOT ratios). Results may be different when the remaining 10 words are rated. If that is the case and the results of ratings align with those of VOT ratios, then these results would be in line with findings by Lord (2005), whose L2 Spanish learners already produced /t/ and /k/ at a native-like level at pre-test. As both the subjects in Lord's and the present study were at the intermediate proficiency level, it may be that voiceless plosives are no longer a pronunciation challenge at the intermediate level, and future studies could investigate the pronunciation features [i] for which practice with a stereotypical L2 accent in the L1 would be beneficial at this level. If the results of ratings do not align with those of VOT ratios, and the Authentic and the Stereotypical groups did fail to improve their pronunciation of the French voiceless plosives, future studies may yield different results with a focus on fewer pronunciation features. Indeed, in a follow-up survey, one participant confided having difficulty processing several pronunciation features at once, while all the others mentioned being grateful for the notes they had taken during training.

While results do show an effect of time for VOT ratios being significantly higher at the delayed post-test than before treatment for the experimental subjects, NS rating results do not. The VOT ratio results could be construed as negative long-term impact of treatment. However, the absence of a significant interaction between time and group indicates that learners still performed at a native-like level one week after treatment. The experimental groups' higher VOT ratios at the delayed post-test could be a case of backsliding (Beebe, 1988), caused by restructuring (McLaughlin, 1987) of the phonetic space as other elements, such as French /ʁ/, vowel stability, and intonation were incorporated. In that case, one week for a delayed post-test may be too early to effectively reflect long-term processing of sounds. Furthermore, as mentioned, the focus on several features at a time may have led learners to information overload, thereby increasing potential for momentary processing confusion. Future studies should arrange for more time between immediate and delayed post-tests, and limit instruction and practice to one feature at a time as in Everitt (2015).

To investigate the relation between VOT ratio measures and native speaker judgements, correlations were run. Significant negative correlations between VOT ratios and NS ratings indicate that as aspiration of /p/ and /t/ decreased towards native-like production, learner production was also rated more native-like. While this was true for both voiceless plosives followed by /a/, it was not for /t/ when preceding /u/, probably due to friction resulting from relaxed muscle tension and a lengthened vocal tract in anticipation of the /u/. Interestingly, as results for the native speaker controls indicate (Table 5), the level of perceived friction seemed to negatively affect the judgement of native speakers, leading them to give /t/ + /u/ lower ratings even when the VOT ratio was low.

Table 5

Native speaker control data for /t/ + /u/

NS Control	NS rating	VOT ration	Level of perceived friction
1	7.00	0.28	high
2	9.00	0.76	low

As explained, these results are limited to 30% of the data and caution must be used when interpreting them. However, they suggest that, at least for voiceless plosives, both VOT measures and rater judgements—informing both on the duration of aspiration and the degree of muscle tension—are necessary for a comprehensive diagnostic of pronunciation quality.

ABOUT THE AUTHOR

Viviane Ruellot is an Associate Professor of French at Western Michigan University, where she teaches French, French linguistics and applied linguistics. Her research interests include feedback on pronunciation, the stages of pronunciation acquisition, computer assisted pronunciation teaching and learning, and the history of French pronunciation teaching.

Contact information:
Western Michigan University
Department of World Languages and Literatures
821 Sprau Tower
1903 W. Michigan Ave.
Kalamazoo, MI 49008-5538
(269)387-3043
viviane.ruellot@wmich.edu

ACKNOWLEDGEMENTS

Many thanks to Cinthia Medina Mendoza and Jesse Klaus, my research assistants, and to three native speaker raters: Véronique Jewell, Sylvie Lidolf, and Violène Pillar. This work was supported by grants from the College Dissemination and Discovery Award and the Support for Faculty Scholars Award, Western Michigan University.

REFERENCES

Beebe, L. M. (1988). The sociolinguistic perspective. In L. M. Beebe (Ed.), *Issues in second language acquisition: Multiple perspectives*. New York: Newbury House.

Boersma, P., & Weenink, D. (2018). Praat: doing phonetics by computer (Version 6.0. 37).

Boucher, V. J. (2002). Timing relations in speech and the identification of voice-onset times: A stable perceptual boundary for voicing categories across speaking rates. *Perception & Psychophysics*, *64*(1), 121-130.

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

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.

Everitt, C. (2015). *Accent Imitation on the L1 as a task to improve L2 pronunciation*. Unpublished MA thesis. Barcelona: University of Barcelona.

Flege, J. E., & Eefting, W. (1987). Cross-language switching in stop consonant perception and production by Dutch speakers of English. *Speech Communication, 6,* 185–202.

Flege, J. E., & Hammond, R. M. (1982). Mimicry of non-distinctive phonetic differences between language varieties. *Studies in Second Language Acquisition*, *5*(1), 1-17.

Flege, J. E., & Munro, M. J. (1994). The word unit in second language speech production and perception. *Studies in Second Language Acquisition*, *16*(4), 381-411.

Honey, J. (2017). Sociophonology. In F. Coulmas (Ed.), *The handbook of sociolinguistics* (pp. 92-106). Malden, MA: Blackwell Publishing

Kelly, L. G. (1969). *25 centuries of language teaching: An inquiry into the science, art, and development of language teaching methodology, 500 BC-1969*. New York: Newbury House Publishers.

Kristiansen, G. (2003). How to do things with allophones: Linguistic stereotypes as cognitive reference points in social cognition. In R. Dirven, R. Frank, & M. Putz (Eds.), *Cognitive models in language and thought: Ideologies, Metaphors, and Meanings* (pp. 69-120). Berlin: Mouton de Gruyter.

Lippi-Green, R. (1997). *English with an accent: Language, ideology, and discrimination in the United States*. London: Psychology Press.

Lisker, L., & Abramson, A.S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word,* 20, 384–422.

Lord, G. (2005). (How) can we teach foreign language pronunciation? On the effects of a Spanish phonetics course. *Hispania*, 557-567.

Major, R. C. (1987). English voiceless stop production by speakers of Brazilian Portuguese. *Journal of Phonetics, 15,* 197–202.

Maxwell, E. M., & Weismer, G. (1982). The contribution of phonological, acoustic, and perceptual techniques to the characterization of a misarticulating child's voice contrast for stops. *Applied Psycholinguistics*, *3*(1), 29-43.

McLaughlin, B. (1987). *Theories of Second-Language Learning*, London: Arnold.

Mora, J. C., Rochdi, Y., & Kivistö-de Souza, H. (2014). Mimicking accented speech as L2 phonological awareness. *Language Awareness*, *23*(1-2), 57-75.

Nagell, K., Olguin, R. S., & Tomasello, M. (1993). Processes of social learning in the tool use of chimpanzees (Pan troglodytes) and human children (Homo sapiens). *Journal of Comparative Psychology*, *107*(2), 174.

Neuhauser, S. (2011). Foreign accent imitation and variation of VOT and voicing in plosives. In *Proceedings of the XVIIth International Congress of Phonetic Science, Hong Kong* (pp. 1462-1465).

Riney, T., & Takagi, N. (1999). Global foreign accent and voice onset time among Japanese EFL speakers. *Language Learning*, *49*(2), 275–302.

Rojczyk, A. (2012). Phonetic and phonological mode in second-language speech: VOT imitation. In *EuroSLA22–22nd Annual Conference of the European Second Language Association, Poznań, Poland* (pp. 5-8).

Rojczyk, A., Porzuczek, A., & Bergier, M. (2013). Immediate and distracted imitation in second-language speech: Unreleased plosives in English. *Research in Language*, *11*(1), 3-18.

Ruellot, V. (2018). *Can Pepé Le Pew help? Stereotypical accent and French pronunciation learning.* 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. 163-173). Ames, IA: Iowa State University.

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*(2), 129-158.

Schmidt, R. W. (1993). Awareness and second language acquisition. *Annual Review of Applied Linguistics*, *13*, 206-226.

Simonds, R. (Producer), & Levy, S. (Director). (2006). *The Pink Panther* [Motion Picture]. United States: Metro-Goldwyn-Mayer.

Summerfield, Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 1074–1095.

Sypiańska, J., & Olender, A. (2016). *Phonetic awareness in Polish learners of English* (pp. 1-9). Poznań Working Papers in Linguistics, *1*(1).

Tanner, M., & Landon, 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.

Thomson, R. I., & Derwing, T. M. (2014). The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics*, *36*(3), 326-344.

[i] Including French /ʁ/, as studied in Ruellot (2018), with a larger number of participants and an exclusive focus on the feature.

Chan, M. J. (2019). Teaching Tip: Smother new or the say mold story? Coke sing Emma cross a notion. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Ames, IA, September 2018 (pp. 354-360). Ames, IA: Iowa State University.

TEACHING TIP

SMOTHER NEWS OR THE SAY MOLD STORY? COKE SING EMMA CROSS A NOTION

Marsha J. Chan, Sunburst Media

Wailing's speech lacks fluency and sounds choppy. She often pronounces words and word segments separately. In particular, she doesn't articulate final /m/ sounds, and she doesn't link final consonants to subsequent vowels in phrases. She is unaware of this /m/ deficit and disconnected speech, yet she is eager to learn to speak English better. With gradual and systematic scaffolding, using techniques that draw upon Wailing's existing skills and tendencies, and supplementing with visual and audio aids, including video clips, this multi-step lesson brings greater perception and production; it helps in coaxing the /m/ across an ocean between words.

Written words are separated on the page by spaces; spoken words in the stream of speech are not. Beginning and intermediate learners are often unable to parse utterances in spoken English, which are typically emitted in continuous streams of sounds carried by prosodic elements of stress, intonation, and rhythm. For example, consonant sounds that belong orthographically to the ends of words are linked to the beginnings of words whose orthographic forms begin with vowels, making the phrase "same old" sound like "say mold" and the phrase "an ocean" sound like "a notion." Linking, as defined by Alameen and Levis, is a category of connected speech processes "that does not involve changes to the segments of the words" but makes two words sound like one… "as in… some_of [sʌm əv]" (2015, p. 162). The importance of linking in pronunciation and listening instruction is widely accepted (e.g., Cauldwell, 2013; Field, 2013; Rost, 2011), and many pronunciation learning materials provide presentations, strategies, and exercises on linking (e.g., Brown 2012; Celce-Murcia, et al., 2010; Chan, 2009; Dauer, 1993; Gilbert, 2005; Grant, 2009; Sardegna & McGregor, 2017). Consonant-to-vowel (C-V) linking is among the most common types of connected speech, yet many adult English learners, especially those who have learned the language through the written form, lack the facility to perceive and produce linked words in speech with accuracy, ease, or confidence.

Some other news or the same old story? Coaxing M across an ocean

This teaching tip focuses on helping learners for whom final /m/ is difficult to produce due to the fact that their primary languages do not use the bilabial voiced continuant at all, or in the same way as in English, even though initial /m/ may be present in their languages. Among them are speakers of Spanish and many dialects of Chinese. Teachers can guide such learners to transfer their success in pronouncing initial /m/ to pronouncing final /m/ in sentences with C-V linking by:

1) describing the place, manner, and voicing characteristics of /m/
2) providing training in lip sensitivity and visual, tactile, and auditory perception
3) leading practice of words with initial, medial, and final /m/
4) guiding production of phrases and sentences with /m/ + vowel
5) using systematic facial expressions and gestures

6) moving an “M” wand to symbolize vibration and linking across word boundaries
7) guiding repetitions of phrase build-up to sentences
8) encouraging the use of video recorded practice material
9) having the learners write and practice their own sentences that include linking /m/ to the next word, after which they record and/or perform them live.

Place, voice, and manner

This sound is bilabial; it is produced with the two lips pressed lightly against each other, with no gap. It is voiced; the vocal cords vibrate. It is a continuant; unlike /b/, which is also bilabial and voiced, the lips part for /b/ as a stop sound, whereas the vibration of the vocal cords continues throughout phonation of /m/. With the lips closed, the sound /m/ is nasal, not oral.

Visual, tactile, and auditory perception

Teachers can train students to increase their sensory perception by having them focus on specific aspects of articulation. Visual: Students look at the teacher’s lips. They then look at their own lips in a mirror and ensure that the lips are touching each other. Tactile: Students place their fingers on their throats to feel the vibration of their vocal cords. They continue feeling the resonance from their throats to their lips. They place a finger on the side of their noses to feel the vibration on the nasal orifice. They move the finger over both nostrils to stop the sound, and then remove the finger to allow continuation of the /m/ sound. Auditory: Students listen to the sound of the teacher or other voice model. They listen closely to their own production of the /m/ to approximate the target sound. A video demonstration can be viewed at Pronunciation characteristics of the sound /m/ (1:28).

Pronouncing /m/ in initial, medial, and final positions

Start with the familiar initial /m/ by having students practice words like *make, more, money, moon, mat, motion, magnificent*, using words that are within their vocabulary range.

Continue with words with medial intervocalic /m/, such as *famous, woman, dreamer, image, camera, amazing, lemon, summer, tomato, command, remain, semester.* Use words in which /m/ is at the beginning of both stressed *(tomato)* and unstressed syllables *(woman)*.

Next, practice words with final /m/, e.g., *come, same, time, mom, some, ham, room, name, team, swim.* If final /m/ does not occur in the students’ dominant languages, they are likely to need more time with these words.

Students may confuse final /m/ with other nasal consonants, such as /n/ and /ŋ/, as well as nasal vowels such as /ã/ or /õ/. If so, you may give the following tips:

- Keep the back of your tongue down; don’t let it close the air passage at the back of the throat to /ŋ/ sound.
- Keep the front of your tongue down; don’t let the tip or the blade approach the teeth or the roof of the mouth, or else it sounds like /n/.

A video demonstration can be viewed at [Pronouncing /m/ at the end of words: come same time mom]() (0:38).

Reconceptualizing boundaries between final /m/ to vowel-initial words

When the students have succeeded in pronouncing initial /m/ and intervocalic /m/, it's time to show them how a final-/m/ word preceding a vowel-initial word can be conceptually reconfigured. Students trained on written text and dependent on seeing letters on a page may tend to make an oral gap between words–as they see a space between written words–rather than link them in spoken English. Sliding the /m/ over to the right and visualizing the letter 'm' in a new position can coax them to pronounce the /m/ connected to the following vowel. In a few cases, real written words and phrases are available to demonstrate the concept for our text-dependent learners.

many
so many
some antics → semantics
some arise → summarize
sam 'n' i → salmon eye
time 'n, again → tie men again
name or money nay more money
summon Ed's friends some o' Ned's friends

Show students that in written English seen in cartoons, dialogs in literature, and movie scripts, an apostrophe is used to replace an omitted or obscured consonant or vowel sound, as above, as well as in these common informally written alterations below:

come on →c'mon
some more →s'more (the name of a dessert consisting of toasted marshmallow and pieces of chocolate bar sandwiched between two graham crackers)

By now, students are prepared to accept respellings, such as the following (choose your variant!), for the purpose of pronunciation:

some apples →se mapples, sa mapples, so**m**apples, s'mapples
some ants →se mants, sa mants, so**m**ants, s'mants
some eggs →se meggs, sa meggs, so**m**eggs, s'meggs

Use phrases that are appropriate for your particular students, their language proficiency, majors, professions, or interests.

Facial expressions, gestures, and written marks

During face-to-face oral practice, you may use systematic facial expressions and gestures to guide students to articulate /m/ more clearly. For example, hum with your lips closed (but not pressed too tightly), a finger pointing to your lips. Place a hand on your throat to emphasize that the vibration needs to start in the vocal cords, move it to a cheek and the nose to indicate that the

vibration continues throughout the oral and nasal cavities. Sometimes, in order to emphasize the tactile and auditory, rather than visual, nature of the vibratory characteristics, I close my eyes while facing my students and humming.
Using phrases with final /m/ preceding a vowel, write phrases with a linking mark such as *farm‿animals, cream‿n‿sugar, some‿other news, same‿old.*

Gesture linking with a flow of your hand or a finger, as in the written form, to indicate the connected speech element as you speak. A video demonstration can be viewed at Linking /m/ in phrases: farm animals, cream ‘n’ sugar, some other news (0:21).

Using an M wand for coke sing Emma cross a notion

For a visual aid, write a big **m** on a card, the size of which is easily visible to your audience. To give a nonverbal correction to an individual while speaking, or to give continued guidance on the target sound to the whole class during practice exercises, simply hold up the **m** card as a reminder of the target sound.

For greater dramatic effect, create an "**m** wand". With a wide marker pen, write a big, bold **m** on a card; affix the card to a popsicle stick or a ruler. Holding the stick, shake the **m** from side to side quickly and in small movements to indicate vibration.

Figure 1. M on a stick.

To coax the /m/ across word boundaries, move the wand smoothly from your right to your left (students see the motion from their left to their right) to emphasize linking of final /m/ during phrases such as *ham‿n‿cheese, Mom‿or Dad.*

Phrase by phrase guided repetitions

Using vocabulary and sentence structures that are appropriate for your particular students, include selected phrases in complete sentences, modeling appropriate phrasing, stress, and intonation. Lead guided repetitions of each phrase, especially the phrases featuring linking of the target /m/ and other sounds, to help them overcome most challenging parts of the sentence. Following repetitions of your model, have students say each sentence on their own–with the text, without the text, to you, to their classmates, softly, loudly, eyes open, eyes closed, sitting down, standing up–to help them build accuracy and fluency.

Examples

Kim is eating a ham and cheese sandwich:

Kim‿is‿eating /
a ham‿an’ cheese /
a ham‿an’ cheese sandwich /
Kim‿is‿eating a ham‿an’ cheese sandwich.

Swimming is an awesome activity:

Swimming /
is‿an‿awesome /
awesome‿activity /
is‿an‿awesome‿activity /
Swimming‿ is‿an‿awesome‿activity!

Tom and you will team up in the same environment:

Tom‿an’‿you /
team‿up /
will team‿up /
same‿environment /
in the same‿environment /
Tom‿an’‿you will team‿up /
in the same‿environment /
Tom‿an’‿you will team‿up in the same‿environment.

A video demonstration can be viewed at Linking /m/ in sentences: Kim is eating a ham 'n' cheese sandwich. Tom 'n' you will team up... (1:47).

Provide recorded practice material

Students benefit from directed practice outside the classroom. Record phrases, sentences, and stories on audio for listening and pronunciation practice. Better yet, especially for the text-bound students who have developed less than optimal /m/ habits, provide video recordings.

If you do not have the time or inclination to create your own, direct your students to access my video lesson on this topic: /m/ the same age the same afternoon the same environment (voiced bilabial continuant) (32 minutes)

Students make, monitor, and record

Having students write down phrases that require /m/ linking and that they commonly use in their daily English-speaking lives–at work, at school, at home, in the community–will help bridge the gap between other people's phrases and their own. Those who are capable can be encouraged to write complete sentences, including ones that include more than one instance of /m/ linking. They may need guidance from you in pronouncing the phrases and sentences, modeling correct pronunciation, stress, intonation, rhythm, and linking, before gaining confidence and fluency in "performing" them live to you or a class, or committing them to an audio or video recording. This part of the process, along with monitoring their output, provides students with a step toward transferring /m/ linking to their free speech.

Teachers of pronunciation, I hope you find these visual and kinesthetic "tricks" help your learners link final /m/ to words beginning with vowels.

ABOUT THE AUTHOR

Marsha J. Chan is an English language specialist, educational consultant, owner of a small business, Sunburst Media, and Professor Emerita of English as a Second Language at Mission College, Santa Clara, California. She has presented over 300 professional seminars and workshops at regional, national, and international conferences and at educational institutions in the USA and abroad. She is the recipient of numerous excellence awards. She is co-founder of CATESOL's Teaching of Pronunciation Interest Group (TOP-IG) and former officer of TESOL's Speech Pronunciation and Listening Interest Section (SPLIS). Author of several English language textbooks, she has created thousands of learning objects in print, audio, and video formats. As Pronunciation Doctor, she provides over 2000 free instructional videos at http://www.youtube.com/PronunciationDoctor.

REFERENCES

Alameen, G., & Levis, J. M. (2015). In M. Reed & J. Levis (Eds.), *The handbook of English pronunciation.* West Sussex, UK: Wiley Blackwell.

Brown, J. D. (2012). *New ways in teaching connected speech*. Alexandria, VA: TESOL International.

Cauldwell, R. (2013). *Phonology for listening: Teaching the stream of speech.* Birmingham, AL: speechinaction.

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

Chan, M. (2009). *Phrase by phrase pronunciation and listening in American English*. Sunnyvale, CA: Sunburst Media.

Dauer, R. M. (1992). *Accurate English: A complete course in pronunciation*. Upper Saddle River, NJ: Prentice Hall.

Field, J. (2003). Promoting perception: Lexical segmentation in L2 listening. *ELT Journal, 57*, 325–334.

Gilbert, J. B. (2005). *Clear speech: Pronunciation and listening comprehension in North American English. Teachers' resource book.* New York: Cambridge University Press.

Grant, L. (2009). *Well said: Pronunciation for clear communication*. (3rd ed.). Boston, MA: Heinle & Heinle Publishers.

Rost, M. (2011). *Teaching and researching listening*. (2nd ed.). Edinburgh Gate: Pearson Education Limited, Longman.

Sardegna, V. G., & McGregor, A. (2017). Oral communication for international graduate students and teaching assistants. In J. Murphy, J. (Ed.), *Teaching the pronunciation of English: Focus on whole courses* (pp. 130-154). Ann Arbor: University of Michigan Press.

Larsson, M., & Boldt, H. (2019). Teaching Tip: Improving intelligibility with prosodic models. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Ames, IA, September 2018 (pp. 361-370). Ames, IA: Iowa State University.

TEACHING TIP

IMPROVING INTELLIGIBILITY WITH PROSODIC MODELS

Margareta Larsson, Georgia State University
Heather Boldt, Emory University

Intonation is "long thought to be a key to effectiveness in spoken language" (Levis & Pickering, 2004). However, producing the expected pattern of intonation presents challenges for students, many of whom find that their production of English suprasegmentals is influenced by their L1, which can lead to obscured meaning. Fortunately, videos of the 3-Minute-Thesis competition (3MT®), held annually at hundreds of universities in over sixty countries, offer high-quality samples of clearly delivered presentations that students can use to model effective intonation. Each video is a short yet first-rate example of English filled with powerful template sentences "in which all levels of the prosodic system are present" (Gilbert, 2014, p. 130). This paper will demonstrate how two experienced practitioners have used 3MT videos to improve the intelligibility of intermediate-level university students. Current best practices covered in this paper include a focus on suprasegmentals, the analysis of authentic language, the importance of target language perception, and the use of gesture to enhance communication. Also, since 3MT presenters are often NNS of English, they serve as valuable "aspirational models" (Murphy, 2014). Though the focus here is primarily on 3MT talks as pronunciation models, our framework can be easily adapted to other models.

INTRODUCTION

Intonation plays "a significant role in communicating to others how we want to be understood" (Pickering, 2018, p. 71). In addition, an increasing amount of research demonstrates that explicit instruction in intonation and other suprasegmental features of language can lead to significant gains in NNS intelligibility, or the extent to which a speaker can be understood (Grant, 2014). As practitioners, then, it behooves us to focus on elements of speech that can lead to improved intelligibility, such as appropriate pausing, syllable stress, prominence, and intonation, in our classrooms. In this paper, we aim to provide practitioners, whether novice or veteran, with engaging, authentic activities that can help intermediate to advanced students improve their intonation. Each of the teaching tips presented is based on research-based best practices and has been refined through many years of classroom application in various academic settings.

The authors taught at the same university for several years and continue to collaborate and present together, both nationally internationally, on the teaching of pronunciation. The tips below resulted from classroom activities related to the Three-Minute Thesis (3MT®) competition, a speaking event founded in 2008 and now held annually at hundreds of universities across the globe (University of Queensland, 2018). The graduate students who most successfully and clearly explain their research to a non-specialist audience win the competition and their videos are often featured on the university's YouTube channel.

Background on Tips 1-3

The first 3 teaching tips were designed for and have been used successfully in an English language support program for NNS graduate students. Using Three Minute Thesis videos as models is a natural fit for this student population since communicating about research is a vital part of their academic lives. The students come from a variety of language backgrounds and tend to be intermediate to advanced low speakers of English on ACTFL's Oral Proficiency of English scale.

Tip #1: Use a broad definition of pronunciation

Segmentals, or the individual sounds of a language, and suprasegmentals, which include stress and intonation, are two main features often associated with pronunciation. However, since an increasing amount of research demonstrates the connection between gesture and clear speech (Goldin-Meadow & Alibali, 2013; Guidetti & Nicoladis, 2008; Smotrova, 2015), instructors should consider broadening their conceptualization of pronunciation to one that includes gesture. In fact, such a framework is provided by Fraser (2001), who describes pronunciation in terms of three key categories: segmentals, suprasegmentals, and peripheral features, such as gesture and other body language. As demonstrated below, instructional activities that involve mirroring videos of winning 3MT presentations will naturally lead to the incorporation of each of these three categories.

Teaching Tip #2: Have students mirror the prosody of high quality models

Mirroring, defined here as shadowing not only the voice characteristics but "the whole bodily action" of the model (Kjellin, 1999, p. 7), has been shown to improve student learning. Designing a mirroring project begins with the selection of videos clips for students to first analyze (for all aspects of pronunciation, as broadly defined above) and then mirror (i.e. reproduce in a performance that mirrors the original speaker as closely as possible). As noted above, winning presentations from the 3-Minute Thesis (3MT®) competition can be an appropriate choice in many academic settings. In addition to being tasked with explaining their research in a way that will interest and engage the audience, 3MT presenters are only allowed to use one static slide so the emphasis is on effective delivery of the spoken word, rather than on visuals. Of course, practitioners in other settings can choose models based on what is most fitting for their contexts. With appropriate models on hand, the next step is to provide students with a structured method for analyzing the speaker's pronunciation. The Prosody Pyramid (Figure 1), provided by Gilbert (2018) and building upon the work by additional scholars, can serve well for this purpose. As can be seen in the figure, a thought group, sometimes also referred to as message unit or breath group, is a series of words linked together with pauses at either end. Within a thought group, the word that receives the most prominence is the focus word, also known as the nucleus, main stress or tonic accent; the default placement of this focus is the last content word in the thought group (Reed & Levis, 2015).

Figure 1. Prosody pyramid (Gilbert, 2014, p. 127).

Within the focus word is the primary stressed syllable, with the vowel in this syllable as the peak of the pyramid.

The concepts above can be practiced in the classroom through working with a 3MT, which is chosen by students based on their interests or by the instructor based on the 3MT's suitability for a particular class. Students listen to a section (no longer than one minute) of the 3MT video (rather than the full three minutes which may prove too time-consuming) and mark a transcript of it for thought groups, focus words, and the stressed "peak" syllables in any multisyllabic focus words. For instance, a student might identify that the phrase "and in our analysis" was linked together, in other words, spoken as one thought group, that the word "analysis" received focus (prominence) from the 3MT speaker, and that the /æ/ sound in the second syllable of "analysis" is the peak vowel. Although the prosody pyramid shows only focuses on the most prominent word in a thought group and its corresponding peak vowel, recent research by Cauldwell shows that in spontaneous speech, about 40% of thought groups have one focus word while an additional 40% of thought groups have two (Murphy, 2017). Applying this concept to the prosody pyramid above, in the thought group */ how do you spell easy /*, "easy" is the strongest focus word and contains the peak vowel but "how" could also receive some focus, leading to a smaller additional peak toward the beginning of the thought group. Although 3MT presentations are not spontaneous speech, this 2-peak pattern appears in many cases, and in the course of analyzing transcripts, students notice the many thought groups that have this two focus word profile.

The sample below shows what a student's analyzed transcript might look like after their initial analysis of thought groups, focus words, and syllable stress within focus words. Thought groups are indicated by the / marks; focus words are marked in bold (students can circle them, highlight them or underline them depending on their preference); the stressed syllables of any multi-syllabic focus words are underlined; peak vowels are written in upper case.

Sample of an analyzed transcript

close your eyes for **tEn** seconds / and **think** about what a **rAInforest** means to you / if you're
like **mE**, / it means **wIld** animals / and **trOpical** plants /**bUt**/ to the **indIgenous** people I
wOrk with / these **forests** mean **fOOd**

Excerpt from Olivia's Sylvester 3MT presentation (University of Manitoba, 2014)

Figure 2. Sample analyzed transcript.

Using their marked transcripts, students can begin to imitate the pausing and intonation patterns of the original speaker. Depending on student level and instructor goals, elements of linking (**think**_about) and reduction (**wild** animals ‘n **tropical** plants) can also be analyzed and imitated. (For an even more detailed description on how to introduce even more analysis and mirroring into your classroom, see teaching tips 5-8 below.) As students analyze their videos and begin to “mirror” the original speaker, they will undoubtedly notice the peripheral features, specifically the gestures, that accompany focus words. In the case of the above excerpt from a 3MT presentation speaker, for instance, the thought group /if you’re like me/ is accompanied with hand movement toward the chest, and in the grouping /these forests mean food/, the speaker points behind her to a picture of the forest. The prevalence of such gestures in effective presentations leads to the next tip.

Teaching Tip #3: Explicitly teach gesture

When mirroring, the salient connection between gesture and focus can lead students to a greater understanding and appreciation for clear delivery. For instance, when mirroring an original speaker producing the thought groups */on the **one** hand/* and */on the **other** hand/,* the use of the hand gesture on the focus words **one** and **other** often helps them deliver these words in the conventional manner. To aid students in their use of gesture, instructors can explicitly teach the three categories of gesture below: representation gestures, pointing gestures, and beat gestures.

1. Representation gestures

   Sometimes referred to as “representational” gestures, this category includes any gesture that represents the word, such as gesturing toward yourself for “my,” making an upward motion with one or both hands for “increase,” or bringing the hands together for the word “collaborate.” Although representation gestures can be further broken down into iconic from metaphoric gestures, grouping them into one category suffices in the classroom.

2. Pointing gestures

   As the name suggests, these gestures involve pointing at or toward a visual, gesturing toward a particular person ("I'd now like to welcome Dean Tedesco") or pointing in a particular direction ("This research is happening right next door").

3. Beat gestures

   Beat gestures refer to hand movements that are not representation or pointing gestures, but instead have the sole purpose of keeping the rhythm. Using a gesture to accompany the peak vowel of a focus word can help speakers produce it with the length and clarity that is expected. For instance, one of the students was mirroring a 3MT presentation that included the sentence "You can see both quantum and relativistic properties at work" (University of Waterloo, 2013), with the focus being on the words "quantum" and "relativistic." The student was having difficulty with production of the focus words, especially the five-syllable word "relativistic," until he added in the original speakers beat gestures (holding up one hand on the peak vowel in **quAntum** and then holding up the opposite hand as he produced the peak vowel in **relativIsti**c). Although anecdotal, such observations are in line with studies on the role of gesture and effective speech (Goldin-Meadow & Alibali, 2013).

After showing students examples of these gesture types, they can add notations to their analyzed 3MT transcript. For instance, they might write "bring hands together" over the word "join" or "point with thumb over shoulder" over the words "in this image." From this point on, when students practice mirroring with their marked transcripts, in class with a partner or at home on their own, they can try to produce focus words together with their corresponding gestures. The vast majority of students, even those who perhaps do not gesture much in their first language, feel quite comfortable incorporating the original speaker's gestures into their mirroring performances. In fact, the act of mirroring is a way for them to "try on" a different way of speaking by imitating a highly intelligible, winning presenter. If the original speaker, however, has a frequency of beat gestures that a student finds hard or awkward to imitate, of course they can choose to simply focus on the representation and pointing gestures, the ones that are perhaps a bit easier to reproduce.

Background on Tips 4-8

These activities are designed for intermediate students in an IEP (Intensive English Program). At first, the whole class works together and analyzes and mirrors one 3MT presentation (Tip #4-7). After that, students follow the same steps and each student works on different 3MT presentations, similar to the mirroring project described above. Finally, each student prepares a poster presentation using a 3MT presentation of their choice (Tip #8).

Teaching Tip #4: Watch a 3MT video without sound and analyze and mirror gestures (whole class activity)

To raise awareness of body language, watch one 3MT video clip selected by the teacher, with the sound muted. To begin, students watch the first 40-45 seconds of the muted 3MT video and guess the words that accompany the gestures. After watching the silent video a few times, students can guess words like "me," "first," and “second," and can also easily recognize phrases, such as "on the one hand," "on the other hand," and “led to a rise in,” simply by interpreting the gestures. Students can then reproduce the clip without sound, practicing and video recording themselves with their smartphones.

Teaching Tip #5: Provide a transcript with punctuation and capitalization removed; mark up the transcript (whole class activity)

To raise awareness of the use of pauses at the boundaries of thought groups, provide a transcript of the selected 3MT presentation without capitalization, commas, and periods and ask the students to mark any pause they hear (Murphy, 2017). Students generally have no trouble indicating the thought group boundaries. This activity helps draw attention to the importance of pausing, and students benefit from incorporating appropriate pausing in their speech because it can lessen the load of the listener and give them time to process speech that might otherwise be challenging to understand. For many NNS student, it is difficult to avoid pausing until the end of a thought group, and this inappropriate pausing within thought groups can cause confusion for the listener.

The next step is to listen and mark focus words, and again, students are generally able to recognize the focus words. Indeed, in many cases, the students have already noted some of these focus words because the 3MT speakers often use gestures to highlight important words (see Tip #4). Students also notice that many of the thought groups have two focus words and that the second word is the strongest (see Tip #2), and mirroring these thought groups helps students produce the multiple peaks and valleys pointed out by Dickerson (2016) in his two-peak model.

After agreeing as a class on the focus words, students are ready to mark intonation using arrows or curved lines. While focusing on the peaks and valleys of intonation, it is important to raise the students' awareness of how native speakers' intonation involves contrasting high and low intonation, and that intonation often falls before it goes back up on a focus word. Although a monotone voice may be intelligible, intonation affects how we perceive other people and how we are perceived. In fact, research at Georgia State University (Clower & Lindeman, 2016), in which undergraduates listened to recordings of the same speakers speaking once in monotone and once with native-like intonation, found that monotone speakers, both native and non-native, were judged as unfriendly, unintelligent, and ineffective while the speakers with appropriate intonation were judged much more favorably.

Teaching Tip #6: Practice in chorus to learn prosody (whole class activity)

Once the transcript is marked up, the teacher leads the students in a chorus practice imitating the original 3MT. Kjellin (1999) promotes chorus practice and intense repetition as a way for language learners to learn prosody, including stress, intonation and rhythm. Just like singing in a chorus, speaking in chorus can help the voice, and subsequently the brain, to acquire the sounds of the new language.

Teaching Tip #7: Students video record themselves mirroring the model 3MT speaker (individual homework assignment)

After analyzing and practicing the sample 3MT together as a whole class, for homework, each student records a video of themselves mirroring about 45-50 seconds of the 3MT presentation. Students seem to put in the most effort in practice when they make videos of themselves, especially if they know that they will have the opportunity to share the videos in class. In fact, most students seem to make more effort when producing a video of themselves than they do preparing for a live presentation of the same material, so this tip can be a powerful tool to encourage sustained practice with intonation.

Teaching Tip #8: Prepare a poster presentation to teach pronunciation using a 3MT (individual project)

Now students are ready to prepare for a poster presentation. The instructor provides transcripts (without punctuation or capitalization, as noted in Tip #5) of the first 45-50 seconds of several 3MTs, choosing models suitable for students' proficiency level, also supplying links to the 3MT videos. Samples include both male and female 3MT speakers and also highly intelligible NNS speaker models, ideally of native speakers from the language backgrounds represented in the class. Examples of accented, NNS communicating effectively can serve as “aspirational models” (Murphy, 2014) so students should be encouraged to use such models.

Next, students follow the steps they learned in the whole-class exercise described above (Tips #4-7). After they mark up the transcript, they mirror the original speaker’s pronunciation and gestures, practicing both in class and for homework. After students have learned to mirror the body language and pronunciation of the 3MT speaker they chose, they prepare to give a poster presentation in which they will teach other students to mirror the presentation. During the poster presentation, students will also teach 2-3 consonant or vowel sounds, choosing the segmentals that they themselves find challenging. In a class with students from different language backgrounds, there will be a wide variety of challenging sounds. For example, a Vietnamese student might focus on consonant clusters, while a Japanese student might focus on /l/ and /r/. In a homogenous class, the students might all focus on similar segmentals, but each poster presentation is based on a different 3MT transcript, so for each presentation the sounds will be taught in a different context.

Finally, each student's poster includes the analyzed portion of the 3MT transcript, marked up with pauses, focus words, and intonation, and also 2-3 specific challenging sounds the student has chosen to teach. Students also illustrate on their posters how the specific consonant or vowel sounds are articulated by drawing images of the mouth and lips. During the poster presentations, each student (1) presents their 3MT, mirroring the body language and the pronunciation of the original speaker; (2) teaches the other students to mirror the 3MT in chorus (teaching gestures, focus and intonation); and (3) teaches sounds that the student finds challenging. Since the format is a poster presentation, each student has the opportunity to present 2-3 times and each presentation takes 4-5 minutes. The other students act as the audience and move around to the 2-3 posters in the room.

The 3MT project described above spans about three weeks from beginning to end. Over several semesters, students have been consistently enthusiastic and engaged throughout this project, and they clearly improve their use of focus words, intonation, and use of gestures in the context of these activities. Since many NNS students often find the stress and intonation patterns of English difficult to acquire and the use of body language awkward, mirroring a model speaker, especially one from their own language background and performing as that speaker, helps lessen the awkwardness. The imitation of the original 3MT speakers is instructive, and another powerful aspect of the project is the students teaching their classmates to mirror the body language and speech patterns of their chosen 3MT speaker. As teachers know, nothing helps learning something as much as having to teach it does. Many students have also shared in their written feedback that the 3MT mirroring project has helped them communicate more clearly beyond the classroom.

CONCLUSION

The teaching tips above are informed by current best practices and center around students' analysis and reproduction of authentic speech using winning 3MT presentations as models. Through analyzing and mirroring the prosody of either NSs or highly intelligible NNSs, students have ample opportunities to practice all aspects of speech, including linking, pausing, syllable stress, prominence, intonation, and gesture. These activities not only raise students' awareness of the prosodic system but also encourage the type of quality repetition that "helps students feel themselves growing in mastery" (Gilbert, 2014, p. 128). As suggested in Tip 8, these mirroring activities can be expanded to include students' teaching each other not only the suprasegmental but also the segmental aspects of the model speaker's 3MT. Students who have engaged in the activities outlined here have consistently reported feeling more confident and more able to communicate effectively, not only in presentations but in other spoken exchanges as well.

ABOUT THE AUTHORS

Margareta Larsson is a Senior Lecturer in the Intensive English Program at Georgia State University in Atlanta, Georgia, USA, where she teaches oral communication, pronunciation, and academic writing. A native of Sweden, she came to the United States as a Fulbright Scholar. She is a graduate of the University of California-Berkeley (MA Anthropology) and Georgia State (MA Applied Linguistics and ESL).

Contact information:
Margareta Larsson
Georgia State University
Intensive English Program
25 Park Place, Suite 1500
Atlanta, GA 30303
404-413-5069

Heather Boldt is an instructor in the English Language Support Program at Emory University in Atlanta, Georgia, USA, where she teaches oral communication and academic writing courses. She is a graduate of Georgia State University (MA Applied Linguistics and TESOL). Her over 20 years

of experience, both in the US and abroad, include teaching in English for Specific Purposes, Intensive English Program, and Adult/Community ESL contexts.

Contact Information:
Heather Boldt
Emory University
English Language Support Program
1784 N. Decatur Road Building, Suite 101
Atlanta, GA 30322
404-727-8075

REFERENCES

Clower, A. & Lindeman, S (2016, April 10). *The role of intonation in attitudes toward non-native speech*. Paper presented at American Association for Applied Linguistics, Orlando, FL.

Dickerson, W. B. (2016). A practitioner's guide to English rhythm: A return to confidence. 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. 39-50). Ames, IA: Iowa State University.

Fraser, H. (2001). *Teaching pronunciation: A handbook for teachers and trainers - Three frameworks for an integrated approach.* Sydney: TAFE NSW, Access Division.

Gilbert, J. (2014). Myth 4: Intonation is hard to teach. In L. Grant (Ed.), *Pronunciation myths: Applying second language research to classroom teaching* (pp. 107-136). Ann Arbor, MI: University of Michigan Press

Gilbert, J. (2018, February). Teaching pronunciation: Simplicity is the Key. *TESOL Connections*. Retrieved from http://newsmanager.commpartners.com/tesolc/downloads/features/2018/2018-02_Pronunciation.pdf

Goldin-Meadow, S., & Alibali, M. W. (2013). Gesture's role in speaking, learning, and creating language. *Annual Review of Psychology*, *64*, 257–283. http://doi.org/10.1146/annurev-psych-113011-143802

Grant, L. (Ed.). (2014). *Pronunciation myths: Applying second language research to classroom teaching*. Ann Arbor, MI: University of Michigan Press

Guidetti, M., & Nicoladis, E. (2008). Gestures and communicative development. *First Language*, *28*(2), 107-115.

Kjellin, O. (1999). Accent addition: Prosody and perception facilitate second language learning. In O. Fujimura, B. D. Joseph, B. Palek (Eds.), *Proceedings of Linguistics and Phonetics Conference 1998*. (pp. 373-398). Prague, Czech Republic: Karolinum Press.

Levis, J., & Pickering, L. (2004). Teaching intonation in discourse using speech visualization technology. *System*, 32, 505-524.

Levis, J. (2018, March). *Cinderella no more! Second language pronunciation research and practice*. Invited presentation at International Teachers of English to Speakers of Other Languages (TESOL) Conference, Chicago, IL.

Murphy, J. (2014) Intelligible, comprehensible, non-native models in ESL/EFL pronunciation teaching. *System*, 42, 258-269.

Murphy, J. (Ed.). (2017). *Teaching the pronunciation of English: Focus on whole courses*. Ann Arbor, MI: University of Michigan Press.

Pickering, L. (2018). *Discourse intonation: A discourse-pragmatic approach to teaching the pronunciation of English*. Ann Arbor: University of Michigan Press.

Reed, M., & Levis, J. (2015). The handbook of English pronunciation. West Sussex, UK: Wiley-Blackwell.

Smotrova, T. (2015). Making pronunciation visible: Gesture in teaching pronunciation. *TESOL Quarterly, 51*(1), 59-89. doi:10.1002/tesq.276

University of Alberta. (2005, May 9). *Hand gestures linked to better speaking* [Public release]. Retrieved February 13, 2018, from https://www.sciencedaily.com/releases/2005/05/050511105253.htm

University of Manitoba (2014, March). *Second Place: Olivia Sylvester, 3MT Final.* YouTube. Retrieved October 31, 2018 from https://www.youtube.com/watch?v=Qm7WoMchZ8I.

University of Queensland (2018, March 20). *About the 3MT*. Retrieved from https://threeminutethesis.uq.edu.au/about

University of Waterloo (2013, April). *2013 University of Waterloo 3MT finalist: Robert Jonsson*. YouTube. Retrieved October 31, 2018 from https://www.youtube.com/watch?v=q0kXeo7OeIE

Lima, E. F., & Zawadzki, Z. (2019). Teaching Tip: Improving speaker intelligibility: Using sitcoms and engaging activities to develop learners' perception and production of word stress in English. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Ames, IA, September 2018 (pp. 371-381). Ames, IA: Iowa State University.

TEACHING TIP

IMPROVING SPEAKER INTELLIGIBILITY: USING SITCOMS AND ENGAGING ACTIVITIES TO DEVELOP LEARNERS' PERCEPTION AND PRODUCTION OF WORD STRESS IN ENGLISH

Edna F. Lima, Ohio University
Zoe Zawadzki, Ohio University

The Teaching Tip described in this paper is designed to help students improve their perception and production of word stress (lexical stress) through technology-enhanced materials. First, students become aware of word stress features through a mini video lecture. Next, they complete a perception activity. In this activity, they watch a short clip from *The Big Bang Theory* and complete a cloze exercise where they choose the word they hear with the correct stress from a dropdown menu (e.g., ex.PER.i.ment or ex.per.i.MENT). Finally, students complete a production activity in which they record themselves imitating the actors using *Audacity*. They record as many times as they wish until they are satisfied with their performance. Once they have finished, they upload their recording to the course site for feedback. The authors also provide suggestions for adapting the focus of the activity (e.g., final intonation patterns), the activity format (e.g., paper-based), and the text genre (e.g., a fable) for authentic listening materials to provide teachers with ideas on how to vary the activity and keep their students engaged and motivated to learn.

TEACHING TIP RATIONALE: THE INTELLIGIBILITY PRINCIPLE

The main goal of the activities presented in this paper is to help improve the intelligibility of adult intermediate learners through perception and production activities focusing on lexical stress (henceforth referred to as word stress) instead of producing native-like speech. Levis (2005) discusses the nativeness versus the intelligibility principles and explains that with the former, the aim is to speak like a native-speaker, whereas the latter focuses on being understood. Intelligibility is the "most fundamental characteristic of successful oral communication" (Derwing & Munro, 2015, p. 1). If an utterance is intelligible, it indicates that the listener can understand what the speaker is trying to convey (Derwing & Munro, 2015). Therefore, intelligibility, not the nativeness principle, should be the goal of pronunciation training.

BACKGROUND

1. The role of suprasegmentals to promote intelligibility

Researchers agree that English suprasegmentals, those features of the language that extend beyond individual sounds, are crucial to L2 speaker comprehensibility and intelligibility and should, therefore, be an integral part of pronunciation instruction (Celce-Murcia, Brinton, Goodwin, & Griner, 2010; Derwing & Munro, 2015). According to Celce-Murcia et al. (2010), errors related to suprasegmentals (i.e., word stress, rhythm, and intonation) may create greater, more serious misunderstandings than segmental (i.e., individual sounds) errors.

When learners understand how suprasegmentals can affect their speech and the meaning of what they say, they are more likely to understand the points of confusion that they produce in a conversation (Gilbert, 2008). When suprasegmental signals are clear, the listener can understand the speaker's message, even if there are errors in segmentals (Gilbert, 2012). Thus, proper control of rhythm, intonation, and word stress are critical for effective communication.

Rhythm is comprised of the combination of stressed and unstressed syllables along with pauses to create a regular, patterned beat in spoken English. When a speaker uses an incorrect rhythm pattern in English, native listeners may not understand what the speaker is trying to convey, or they may grow frustrated (Celce-Murcia et al., 2010). Gilbert (2012) argues that rhythm serves as a signal to help the listener understand the speaker, create emphasis, and identify relationships. Although these signals are not usually apparent to nonnative speakers, they are essential to native listeners; therefore, the improper use of rhythm may lead to conversational breakdowns between nonnative speakers and native listeners as well as between nonnative speakers from a variety of first language (L1) backgrounds.

Intonation, another suprasegmental that can greatly impact speaker intelligibility and comprehensibility, is described as the "rising and falling of the voice to various pitch levels during the articulation of an utterance" (Celce-Murcia et al., 2010, p. 231). When a speaker uses incorrect intonation patterns, miscommunication is likely to occur. For instance, based on their overuse of rising intonation, English learners may often sound uncertain about what they are trying to communicate. Speakers may also be considered "unfriendly or unengaged" because of an "overuse of falling tones accompanied by some level tones" (Pickering, 2018, p. 54). According to Reed (2013), intonation can also indicate turn taking in conversations. Hence, if improperly used, intonation may cause misunderstanding between interlocutors.

As with rhythm and intonation, word stress may also cause communication barriers. Stressed syllables are those syllables that are longer, louder, and higher in pitch; in other words, native speakers emphasize stressed syllables through length, volume, and pitch. Considering the listener's perspective, "the most salient features of stress are probably longer vowel duration in the stressed syllable and higher pitch" (Celce-Murcia et al., 2010, p. 184).

Native speakers also de-emphasize unstressed syllables by reducing the vowel of those syllables (Celce-Murcia et al., 2010). Word stress is so critical in English that "even if all the individual sounds are pronounced correctly, incorrect placement of stress can cause misunderstanding" (Celce-Murcia et al., 2010, p. 198). Levis (2018) argues that "misplaced word stress in English can stop communication completely" (p. 100), as listeners may stop processing information to try to figure out the word that they did not understand.

2. The role of perception training in pronunciation

Perception is an important skill that should be considered when teaching pronunciation. In fact, both perception and production are key to good communication (Dickerson, 2015). According to Levis (2018), "perception is at least equally important to production" (p. 241). He claims that "if an L2 speaker struggles with production, it is likely that their difficulties have roots in perception" (p. 241).

Flege, MacKay, and Meador (1999) found that accuracy of production (of segmentals) is connected to accurate perception of the features. Improvement in oral production has been noted after the completion of perceptual training (Derwing & Munro, 2015). Therefore, learners should receive instruction on perception and engage in perception practice activities before they complete production tasks. The raised awareness of learners in regard to perception ability can improve their production skills (Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1997; Kissling, 2014, Lambacher, Martens, Kakehi, Marasinghe, & Molholt, 2005).

Kissling (2014) noted that the ability to discriminate and perceive sounds from a native speaker of Spanish had a significant impact on the production of the target sounds. Although this study investigated the perception and production of segmental features, there have also been claims that the perception of suprasegmentals can affect production and cause communication issues (Cutler, 2015; Levis, 2018; Rivers, 1981).

Perception needs to be considered in regards to intelligibility, particularly with rhythm, as L2 learners need to not only be understood but also to be able to understand others' speech as well (Levis, 2018; Rivers, 1981). According to Cutler (2015), both the perception and the production of word stress in English can create issues in communication, either directly or indirectly. Consequently, it is critical to teach both production and perception skills when focusing on pronunciation.

3. Technology (CALL) and pronunciation training

As previously mentioned, the activities discussed in this teaching tip include technology-implemented materials and authentic listening texts to keep students engaged and motivated to learn. According to Barger and Byrd (2011), the employment of computers and technology motivates students since they often enjoy using computers for non-academic purposes; Keller and Litchfield (2002, as cited in Barger & Byrd, 2011) explain that motivation occurs at three different levels: motivation to learn, motivation to work, and self-motivation.

Instruction of suprasegmentals through technology has the potential to lead to improved intelligibility; a study by Lima (2015a) showed that participants improved their pronunciation after computer-assisted instruction on word stress, rhythm, and intonation. Several of Lima's (2015a) activities made use of sitcom clips, which provide authentic input and the possibility of more engaging and motivating pronunciation practice for students (Lima & Levis, 2017). Activities including these types of clips allow teachers to use drama and imitation techniques, as they "offer discourse-level practice with stress, intonation, and connected speech" (Goodwin, 2013, p. 7).

Imitation of native speakers "suggests an effect for rehearsal and reflection," which indicates a possible "connection between the phonological loop function and actual phonological learning" (Moyer, 2014, pp. 429-430). In other words, the phonological loop function is related to humans' working memory and "allows the listener to hold and rehearse sound sequences in short-term memory during speech processing, and to direct attention and promote subvocal articulation that feeds into long-term memory" (Moyer, 2014, p. 428).

It is noteworthy that the tasks described in this paper are part of the Supra Tutor, an eight-week fully online pronunciation course designed to help students improve their use of English suprasegmentals and, in turn, their intelligibility. Also, given that the teaching tips presented at the Pronunciation in Second Language Learning and Teaching (PSLLT) conference are brief due to the nature of the presentation, word stress activities were chosen for the demonstration. However, in the last section of this paper, the authors provide suggestions for adaptations of target features (e.g., intonation patterns).

Before completing the word stress perception and production exercises, students are made aware of the features of word stress through a mini video lecture. The lecture (see Task 1 below) includes the characteristics of word stress, provides information on word stress from both speaker and listener perspectives, and shows visual representation of word stress using *Audacity*. Next, they complete a cloze exercise to practice their perception of word stress, followed by an imitation activity for production practice.

Goals of the Teaching Tip

1. To raise awareness of word stress and why it is critical to successful communication.
2. To understand the features of word stress from both speaker and listener perspectives.
3. To develop self-monitoring skills for pronunciation improvement.

HOW TO CONDUCT THE ACTIVITY

Task 1: Video lecture

The first task in this activity prompts students to watch a mini video lecture. Since the course is online, students may watch the video as many times as they wish. However, if implementing this activity in a face-to-face environment, the authors suggest playing the video at least twice. This mini video lecture allows students to become aware of the features of word stress and why it is critical for successful communication. For instance, the lecture seen in Figure 1 describes the characteristics of word stress, discusses word stress from the perspectives of both speaker and listener, and shows visual representations of word stress using *Audacity.*

The mini video lecture can either be created by the instructor, as is the case of the Supra Tutor, or be found online. In either case, it is important to keep content accuracy and length in mind. Short videos (maximum of 15 minutes) are preferable in order to keep students engaged and attentive while watching the lecture.

Figure 1. Word stress lecture used for Task 1.

Task 2: Perception activity

After students have learned about the features of word stress, they complete a perception activity (Figure 2). In this activity, they watch a short clip from *The Big Bang Theory* (Football Questions). They may watch the clip as many times as they wish. The students then complete a cloze exercise where they choose, from a dropdown menu (e.g., FOOT.ball or foot.BALL), the word that they hear with the correct primary stress. The stressed syllable is represented by capital letters, and periods show syllabification, or the division of words into syllables.

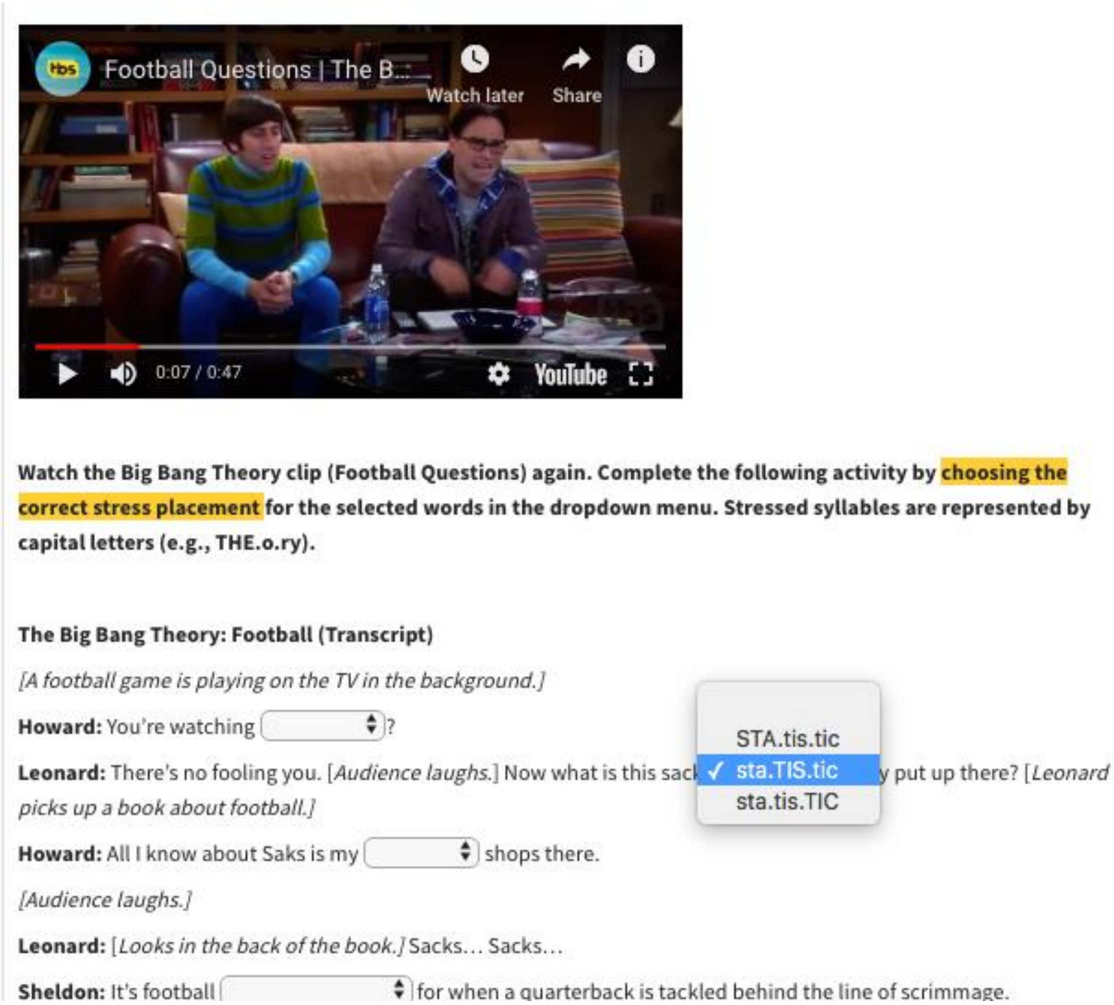


Figure 2. Online word stress perception activity used for Task 2.

While selecting the words for this specific activity, the authors chose content words containing between two and four syllables, given that the content words are naturally emphasized by native speakers. Since the students complete the activity online, it allows for automatic immediate feedback. The system will let students know if the answer they selected is correct or incorrect; if their choice is incorrect, the system will show what the correct answer is.

Although this activity is part of a fully online course, this task can easily be adapted to a paper version. Instead of choosing from a drop-down menu, the students can circle or underline the options showing correct word stress placement. The words from which the students must select may be in bold so that they are easily identifiable. For feedback, the students can either turn in their worksheets to the teacher, or the teacher can go over the correct answers with the class as a whole. A sample paper-based version of the activity is provided below.

The Big Bang Theory: Football (Transcript)

[A football game is playing on the TV in the background.]
Howard: You're watching (**FOOT.ball / foot.BALL**)?
Leonard: There's no fooling you. *[Audience laughs.]* Now what is this sacks (**STA.tis.tic / sta.TIS.tic / sta.tis.TIC**) they put up there? *[Leonard picks up a book about football.]*
Howard: All I know about Saks is my (**MOTH.er / moth.ER**) shops there.
[Audience laughs.]
Leonard: *[Looks in the back of the book.]* Sacks… Sacks…
Sheldon: It's football (**no.MEN.cla.ture / no.men.CLA.ture / no.men.cla.TURE**) for when a quarterback is tackled behind the line of scrimmage.
Leonard: Huh. *[Looks in book again.]* Scrimmage.
Sheldon: The line of scrimmage is the (**IMAG.i.nary / imag.I.nary / imag.i.NARY**) transverse line separating the (**OF.fense / of.FENSE**) from the defense.
Leonard: Oh.
Howard: Sheldon knows football?
Leonard: (**AP.par.ent.ly / ap.PAR.ent.ly / ap.par.ENT.ly**).
Howard: I mean (**QUIDD.itch / quidd.ITCH**), sure, but football?!

Task 3: Production activity

In this last step, students complete a production activity (Figure 3) where they record themselves in *Audacity*, imitating the actors. For this, a segmented audio file is created and provided to the students. By the time this activity is conducted, students will have already learned how to download and use *Audacity* to record and edit their own audio files. In the case of the online course, mini tutorial videos show students, step-by-step, how to download, install, and use Audacity. The audio-clips are segmented into shorter utterances so that the students can easily record themselves right after the actors or, in this case, after the models. Students record themselves as many times as they wish and then listen to themselves for self-monitoring. Once they are satisfied with their performance, the students save the audio file as one single track and upload it to the course website. To learn how to segment audio files using *Audacity*, see Lima (2015b).

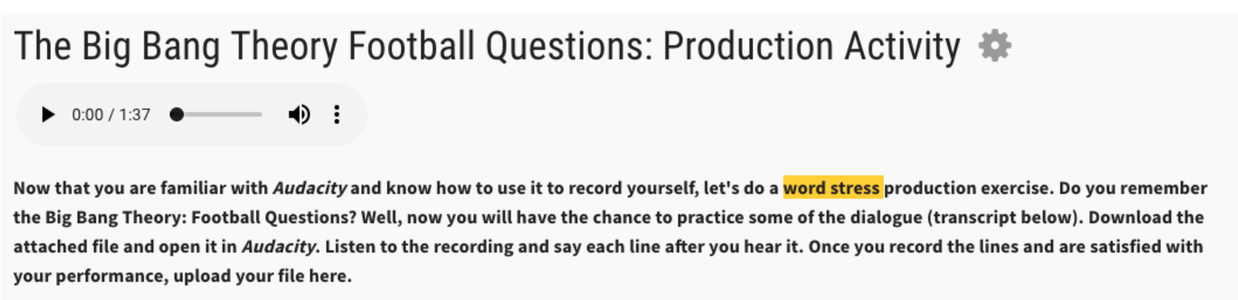


Figure 3. Word stress production activity (Task 3).

Alternatively, instructors can require students to monitor their own progress on the task and submit a short oral (or written) reflection on how they think they did, discussing both their strengths and weaknesses. This will allow for additional spoken practice. In the case that this activity is being performed in class, there are multiple ways in which it can be adapted. For instance, students can repeat after the actors as a group or individually. Additionally, they can role-play the scene in pairs or in groups of three. Considering that self-monitoring is critical to pronunciation improvement (Morley, 1991), students could also record themselves on their cell phones for self-monitoring. The instructor can provide a brief rubric containing key features of word stress to help students focus their self-analyses. Additionally, students could record themselves on computers or on their cell phones and exchange recordings with each other and provide peer feedback using the provided rubric.

ADAPTATIONS OF THE ACTIVITY

The activities described in this teaching tip include technology-enhanced materials (e.g., video lectures and short sitcom clips) to engage and motivate students to learn and to practice. A *Big Bang Theory* clip was chosen for the activities demonstrated here, but any other appealing sitcom can be used (e.g., *Friends*, *Seinfeld*, *Last Man Standing*, *New Girl*, and so on). It is important that instructors choose shows that are appealing to the learners even if they, the instructors, are not very fond of a particular show. The point is to have students interested in the task so that they do not feel discouraged, bored, or unwilling to practice.

Before discussing adaptation options, the authors suggest including a scaffolding activity focusing on accuracy in stress placement at word level first. For instance, the instructor can present the students with a recorded list of words, and the students have to identify (e.g., circle or underline) the syllable that receives the primary stress. Once students master identifying the stressed syllable (perception) and producing it accurately in production tasks, they can move on to phrases and then sentence-level activities such as the one presented in Task 3.

For text variety, the same exercise can be created using short stories or fables (e.g., The boy who cried wolf). Teachers can either use an existing video (e.g., https://bit.ly/1UNnUtH) or record their own versions of the fable. Sometimes the existing recordings are geared towards children, which older learners may find unsuitable or even irritating. Therefore, teachers may find it more appropriate to record their own versions. Fables are also good for teaching thought groups, for example. The students can have the script in front of them and mark the thought groups with slashes (e.g., Once upon a time/ there was a boy/ who had to look after sheep. //) as they hear the story.

As mentioned previously, the activities described above were designed with adult intermediate learners in mind. However, the materials can easily be adapted for a different audience. Depending on the level of proficiency, changes may be needed. For instance, with lower levels, instructors can use shorter clips or texts that are easier for learners to process and understand.

While the demonstrated activities focus on word stress, they can be adapted to focus on other suprasegmental features, such as prominence or intonation patterns. Sitcom clips, cloze exercises, and imitation activities can still be the backbone of the activity; what will change is the focus of the mini-lecture and of both perception and production tasks. For instance, the same clip (*The Big Bang Theory*, Football Questions) can be used for final intonation patterns practice. After watching a mini-lecture video on intonation, students can complete a cloze exercise that prompts them to choose the correct final intonation pattern (either rising or falling). Alternatively, students could be asked to draw arrows next to each utterance according to the final intonation pattern that they hear. For production, the same segmented audio file can be used, but this time learners will focus on imitating the actors in their use of final intonation patterns.

With authentic or semi-authentic texts, the options for perception and production pronunciation practice are limitless. From specific segmentals (e.g., /s/ as in 'Sue' versus /ʃ/ as in 'shoe' or /æ/ as in 'bat' versus /ɛ/ as in 'bet') to intonational discourse, instructors can create fun and engaging activities that focus on any given feature, and students can enjoy pronunciation practice that goes far beyond the traditional pronunciation drills.

ABOUT THE AUTHORS

Dr. Edna Lima is a lecturer in the ELIP Program/Linguistics Department at Ohio University, where she is currently teaching Oral Communication Skills for International Teaching Assistants, Writing for Research to domestic and international graduate students, and Computers in Language Teaching II for MA students in Linguistics. She is also an instructor and developer of Online Certificate Programs. Edna has taught EFL/ESL/ESP for over 20 years. She earned her Ph.D. in Applied Linguistics and Technology from Iowa State University. Her research interests include computer-assisted language learning (CALL), especially in regards to pronunciation instruction, language assessment, and material design.

Contact information:
383 Gordy Hall
Athens, OH 45701
lima@ohio.edu
(740) 593-4538

Zoe Zawadzki is currently a second-year master's student in Applied Linguistics at Ohio University. For the past 16 months, she has been working closely with her mentor, Dr. Edna Lima, to expand and revise an online pronunciation course. Her main research interests include computer assisted language learning (CALL), second language acquisition, and pronunciation. Zoe will be starting her Ph.D. in Applied Linguistics and Technology at Iowa State in the Fall of 2019.

Contact information:
383 Gordy Hall
Athens, OH 45701
zz996711@ohio.edu

REFERENCES

Barger, A., & Byrd, K. (2012). Motivation and computer-based instructional design. *Journal of Cross-Disciplinary Perspectives in Education,* 4(1), 1-9.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America*, *101*(4), 2299–2310.

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

Cutler, A. (2015). Lexical stress in English pronunciation. In M. Reed & J. M. Levis (Eds.), *The handbook of English pronunciation* (pp. 106–124). Malden, MA: Wiley Blackwell.

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

Dickerson, W. (2015). Using orthography to teach pronunciation. In M. Reed & J. Levis (Eds.), *The handbook of English pronunciation* (pp. 488-504). Malden, MA: Wiley Blackwell.

Flege, J. E., MacKay, I. R., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *Journal of the Acoustical Society of America*, *106*(5), 2973-2987.

Gilbert, J. B. (2008). *Teaching pronunciation: Using the prosody pyramid*. New York, NY. Cambridge University Press.

Gilbert, J. B. (2012). *Clear speech: Pronunciation and listening comprehension in North American English* (4th ed.). New York, NY: Cambridge University Press.

Goodwin, J. (2013). Pronunciation teaching methods and techniques. In C.A. Chapelle (Ed.), *The encyclopedia of applied linguistics*. Oxford, UK: Blackwell Publishing Ltd. https://doi.org/10.1002/9781405198431.wbeal0970.

Keller, J., & Litchfield, B. (2002). Motivation and performance. In R. A. Reiser & J. V. Dempsey (Eds.), *Trends and issues in instructional design and technology* (pp. 83-98). Upper Saddle River, NJ: Merril Prentice Hall.

Kissling, E. M. (2014). What predicts the effectiveness of foreign-language pronunciation instruction?: Investigating the role of perception and other individual differences. *The Canadian Modern Language Review*, *70*(4), 532-558.

Lambacher, S. G., Martens, W. L., Kakehi, K., Marasinghe, C. A., & Molholt, G. (2005). The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Applied Psycholinguistics, 26*(2), 227–247.

Levis, J. M. (2005). Changing contexts and shifting paradigms of in pronunciation teaching. *TESOL Quarterly*, *39*(3), 369-377.

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

Lima, E. F. (2015a). Development and evaluation of online pronunciation instruction for international teaching assistants' comprehensibility (Doctoral Dissertation). Available from ProQuest Dissertations and Theses database.

Lima, E. F. (2015b). Feel the rhythm! Fun and effective pronunciation practice using *Audacity* and sitcom scenes (Teaching Tip). In J. Levis, R. Mohamed, M. Qian & Z. Zhou (Eds), *Proceedings of the 6th Pronunciation in Second Language Learning and Teaching Conference* (ISSN 2380-9566), Santa Barbara, CA (pp. 274-281). Ames, IA: Iowa State University.

Lima, E. F., & Levis, J. M. (2017). A university-based online "pronunciation tutor." In *Teaching the pronunciation of English: Focus on whole courses* (pp. 155–175). Ann Arbor, MI: University of Michigan Press.

Morley, J. (1991). The pronunciation component in teaching English to speakers of other languages. *TESOL Quarterly*, *25*(3), 481–520.

Moyer, A. (2014). Exceptional outcomes in L2 phonology: The critical factors of learner engagement and self-regulation. *Applied Linguistics*, *35*(4), 418–440.

Pickering, L. (2018). *Discourse intonation: A discourse-pragmatic approach to teaching the pronunciation of English*. Ann Harbor, MI: University of Michigan Press.

Reed, B. S. (2013). Suprasegmentals: Prosody in conversation. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics*. Oxford, UK: Blackwell Publishing Ltd. DOI: 10.1002/9781405198431.wbeal1311

Rivers, W. M. (1981). *Teaching foreign-language skills*. Chicago, IL: The University of Chicago Press.

Long, A. Y., Alarcon, L., & Ruiz-Pérez, S. (2019). Teaching Tip: Task design for second language Spanish fluency development. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Ames, IA, September 2018 (pp. 382-392). Ames, IA: Iowa State University.

TEACHING TIP

TASK DESIGN FOR SECOND LANGUAGE SPANISH FLUENCY DEVELOPMENT

Avizia Y. Long, San José State University
Lorena Alarcon, Texas Tech University
Sergio Ruiz-Pérez, Texas Tech University

This teaching tip offers a guide for task design aimed at promoting fluency development in intermediate-level second language (L2) Spanish. Tasks—defined as language-teaching activities during which learners negotiate for meaning to achieve a nonlinguistic outcome (e.g., Ellis, 2009)—have been shown to encourage development of L2 grammar and lexis. Recently, scholars have investigated the role of tasks in promoting L2 pronunciation development (see Gurzynski-Weiss, Long, & Solon's [2017] special issue), and this growing body of work has focused on segmental and suprasegmental aspects of L2s. To demonstrate task design for fluency development, we present and explain in detail our task that incorporates stories presented in videos and encourages meaningful practice for intermediate-level classroom learners of Spanish. Our detailed description of the task and its objectives are followed by an evaluation of how it meets the criteria for a task (Ellis, 2009) and suggestions for implementation and manipulation. From this teaching tip, the reader will learn how to design a task to promote fluency in a L2.

INTRODUCTION

This teaching tip outlines the design and implementation of a task aimed to facilitate development of second language (L2) Spanish fluency at the intermediate level. We begin by defining task and offering a brief overview of the theoretical rationale underlying tasks from the perspective of task-based language learning and teaching. We then describe our task, offering a detailed overview of each component of the task as well as guidelines for implementation and manipulation. In addition to the description detailed here, the task can be found at Indiana University's task-based language teaching website (http://tblt.indiana.edu/tasks.html).

What is a task?

A task is the central unit for language-based lesson and program design within a task-based approach to second language teaching. There are several approaches to how tasks may facilitate language development (e.g., Ellis, 2003; Long, 1985; Skehan, 1998); however, each approach shares a focus on authentic language use as well as a focus on form to draw learners' attention to language during communication (Ellis, 2009).

Each task-based approach to language teaching has its own definition of a task that may differ (to a greater or lesser extent) from other approaches. Across the variety of definitions that exist, Ellis (2009) proposed that, for a language activity to be considered a task, it must have the four characteristics outlined in Table 1.

Table 1

Task components (Ellis, 2009)

Task component	Description
Primary focus on meaning	A task has a central focus on communication in the target language.
Negotiation of meaning	Learners interact in the target language to achieve a communicative outcome. Negotiation of meaning may occur through clarification requests, confirmation checks, comprehension checks, etc.
Learners must use their own linguistic and nonlinguistic resources	Learners use their own knowledge of the target language, as well as gestures and their general knowledge of the world, during interaction.
Communicative outcome	The specific goal of the task is a non-linguistic, communicative outcome (e.g., deciding on a restaurant, determining who the thief is, etc.).

Following Ellis's (2009) guidelines, a contextualized fill-in-the-blank exercise in Spanish that requires learners to provide the correct form of *ser* or *estar* would not be considered a task. Writing an ad in which learners describe their ideal roommate, on the other hand, would be considered a task. The task described in this teaching tip adopts Ellis's (2009) criteria for task design.

With respect to task design, tasks are also characterized by whether or not a specific linguistic structure is targeted. Focused tasks are designed to orient learners' attention to a specific linguistic structure, and attention to and/or production of this structure is required in order to complete the task. The focused task in Solon, Long, and Gurzynski-Weiss (2017) required learners to attend to vowels to complete a Spanish language map task by including minimal pair street names (e.g., *Calle P**i**co*, *Calle P**e**co*). Unfocused tasks, on the other hand, do not target a specific linguistic structure. For example, Gilabert's (2007) decision-making task, for which learners had to make a series of decisions as fire chief to save individuals from a burning building, was not designed to elicit a specific linguistic structure. The task described in this teaching tip is an unfocused task; it was not designed to elicit a specific linguistic structure from learners during task completion.

Tasks are also characterized by whether they are designed to provide input or prompt output. As these labels suggest, tasks can be designed to provide receptive exposure to the target language (typically by means of listening or reading) or to encourage learners to produce language (by means of speaking, writing, or integration of a receptive and productive skill). The task described in this teaching tip was designed to prompt output.

Finally, tasks may also be designed to be repeated, that is, repeating a specific task multiple times (cf. procedural repetition; e.g., Kim & Tracy-Ventura, 2013). Research on task repetition has shown that repeating a task has positive effects on interaction (Ellis, 2003) and leads to more effective communication (e.g., Yule, Powers, & McDonald, 1992). Task repetition has also been shown to improve fluency in the target language (e.g., Bygate, 1996). The task described in this teaching tip was designed to be repeated in order to promote improved fluency in Spanish as a L2.

What can tasks do for second language pronunciation?

While tasks have been known to promote second language development of vocabulary, grammar, and pragmatics, the potential benefits of tasks for L2 pronunciation development reflects a fairly recent empirical area (see Gurzynski-Weiss et al. 2017, special issue on the benefits of tasks for L2 pronunciation). Tasks have been shown to be (at least partially) beneficial for L2 development of segmental phenomena (e.g., Solon et al., 2017) and suprasegmental phenomena (e.g., Jung, Kim, & Murphy, 2017; McKinnon, 2017; Parlak & Ziegler, 2017). Tasks have also been shown to direct learners' attention to L2 pronunciation by encouraging explicit discussion of pronunciation during task completion (e.g., Solon et al., 2017; Loewen & Isbell, 2017).

Of direct relevance to this teaching tip, tasks have been shown to facilitate fluency development by means of task repetition (e.g., Bygate, 1996). A study by Lambert, Kormos, and Minn (2017) demonstrated that, regardless of learner level and the type of task being repeated (e.g., narration, opinion), learners showed gains in oral fluency as a result of repeating a task. Fluency is tied to other pronunciation-related constructs such as comprehensibility. While comprehensibility is a multilayered construct, research has shown that speech rate (one way in which fluency has been operationalized in L2 production research), amongst other aspects of L2 speech (e.g., vocabulary and prosody), influence comprehensibility ratings for beginning and intermediate-level learners (Saito, Trofimovich, & Isaacs, 2016). Further, fluency has been shown to negatively impact comprehensibility of Spanish spoken by native English-speaking learners (McBride, 2015), who represent the target learners of this teaching tip. Taken together, it stands to reason that task repetition (which is a feature of task design) may facilitate fluency in a L2, which in turn may lead to improved comprehensibility in the L2. This teaching tip outlines basic elements of task design (that includes task repetition) to target L2 fluency development.

THE TASK: *EL ROBO DEL DIAMANTE* 'THE DIAMOND ROBBERY'

Overview

The task featured in this teaching tip was designed for learners of Spanish in their second or third year of college-level courses. This two-way information-gap task requires learners to work in pairs, and the goal of the task is to piece together the main events and details surrounding the mysterious disappearance of a scientist. Although the *communicative outcome* is the collaborative retelling of the story, the *intended aim* of the task is to encourage improvement in fluency.

The story is presented in two different videos: *Origen del diamante* (see Figure 1) and *Reporte: La desaparición de César Cabral* (see Figure 2). These videos were created with the free animation software Powtoon and scripted and voiced over by the authors of this teaching tip. Students in each pair are assigned to watch one of the videos at home. Students also have the option to individually watch their assigned videos again before doing the task by using their own technological devices. Hence, student A watches the video *Origen del diamante* and Student B watches *Reporte: La desaparición de César Cabral*. For the task, Student A and Student B are instructed to determine the main events of the story and reach a decision about what happened to the main character, César Cabral, and the diamond he created. After students in pairs reach a decision, they are instructed to watch their assigned videos again and repeat the task of determining the main events of the story and of deciding what happened to César Cabral and his diamond.

Figure 1. Screenshot of Origen del diamante.

Figure 2. Screenshot of *Reporte:* La desaparición de César Cabral.

The story presents César Cabral, a scientist from Madrid, Spain, who managed to transform carbon into a diamond at his laboratory in Argentina. The main reason why he worked so hard toward this goal was to please his mother, who always dreamt of owning this kind of jewel but could never afford one. Over several months, César spent many days and nights in the laboratory he shared with two other scientists, Florencia and Hermenegilda, working on his project to the point that his obsession turned into an issue and Florencia decided to take a trip to Ecuador in order to spend time away from the laboratory. The story also reveals that César described his project to some of his acquaintances (e.g., his gym coach, his barber, etc.), after which he begins receiving anonymous threats. His suspicion of everyone motivates his decision to flee, and he leaves a recorded confession (*Origen del diamante*) and a diamond behind. The mystery is where César is and who is threatening him.

Is *El Robo del Diamante* a task?

Recall that a task minimally includes the design components outlined in Table 1. In this section, we address how each of the task components are addressed in our task.

The first component of a task is that there is a focus on meaning. Our task meets these criteria by requiring learners to attend to communication (as opposed to metalinguistic aspects of the L2) during task completion. Additionally, to incorporate authentic linguistic variation in the materials, the *Origen del diamante* video was recorded by one of the authors who is from Spain whereas the *Reporte: La desaparición de César Cabral* video includes the voice of one of the other authors who is from Argentina. The second component of a task is that learners are required to negotiate for meaning. Our task meets this criteria by including a gap in information. The information gap of this task lies in the fact that each video contains some information that the other video does not. Our task also meets the negotiation-for-meaning criteria in that the perspective is different for Student A and Student B: Student A (*Origen del diamante*) watches a first person narrative by the main character of the story whereas Student B (*Reporte: La desaparición de César Cabral*) watches a third person narrative by a TV news reporter.

The third component of a task is that learners must use their own linguistic and non-linguistic resources. In our task, Students A and B need to use their own resources to communicate the events and to discover the missing information during interaction. To illustrate, in the video *Origen del diamante*, César Cabral wonders if his coworker Florencia took a break because of his intense work at the laboratory, but he is not aware that she is back in town, a fact which is explicitly presented in *Reporte: La desaparición de César Cabral*. Student A and B must use their knowledge of the L2 to communicate in the language, as well as general cognitive processes (e.g., selection, classification, reasoning, evaluation, etc.; see Ellis, 2003) to complete the task. As suggested by Ellis, processes such as selection, classification, etc. influence the selection of language without determining it. These processes also limit the range of linguistic forms that a user will need to complete the task.

Last, a task must have a nonlinguistic, communicative outcome. At no time during our task are students instructed to use specific forms; they are simply instructed to determine the main events of the story and reach a decision about what happened to César Cabral (the main character of the story) and his diamond. Table 2 offers a concise summary of how our task meets the criteria for task design.

Table 2

Task components in teaching tip task

Task Component	Description	Teacher Tip Task Component
Primary focus on meaning	A task has a central focus on communication. Although the task can also be focused on form, the emphasis on meaning suggests that communication is a central factor.	Student A and Student B narrate to each other the events they recall from *Origen del diamante* and *La desaparición de César Cabral*, respectively.
Negotiation of meaning	Learners interact and negotiate meaning to complete the task. Negotiation of meaning can occur for example through clarification requests, confirmation checks, comprehension checks, etc.	Both students collaboratively reconstruct the story of *El robo del diamante* using distinct information from their assigned videos.
Learners must use their own linguistic and nonlinguistic resources	Tasks suggest that learners can use their own resources to communicate meaning. An example of linguistic resources includes the use of a particular form. Other example of nonlinguistic resources includes gestures learners might use in the interaction.	Students use their own knowledge of the language (e.g., forms such as the simple present, present progressive, preterite/imperfect, etc. and vocabulary related to professions, locations and materials) to describe what César and other story characters or the news reporter said and did. Students use general resources such as reasoning and evaluation to complete the task.
Communicative outcome	Tasks have specific goals, related to communication, which indicate that learners have completed the task successfully.	Both students come to an agreement on what happened, as well as the whereabouts of César Cabral and his diamond.

Recall that a task may further be designed as a focused or unfocused task, as well as an input-providing or output-prompting task. Our task was designed as an unfocused task, as no specific linguistic structure(s) of the L2 are required to complete the task. The task could be manipulated to target a specific pronunciation feature. For example, the names of the story's characters could be altered to encourage attention to vowels (e.g., Roberto for the main character, Roberta for his lab assistant, etc.), which are important for pronunciation in L2 Spanish. Our task was also designed as an output-prompting task, given that the purpose of interaction is to encourage production in the L2. Again, the task could be manipulated to provide input in the L2. For instance, if students were learning new vocabulary related to places (e.g., *biblioteca, laboratorio*, etc.), the goal of the task could be to determine all of the places that the story characters frequented. During task completion, Student A and Student B would be instructed to determine each location visited by the story's characters based on their individually assigned videos.

Learning objectives and implementation

The specific learning objectives of this task are as follows: Students will be able to:

- **identify** the main events and details of their assigned story depicted in a video;
- collaboratively **summarize** the main events and details of a story depicted across two videos;
- collaboratively **discuss** the chronological order of the main events and details presented in the videos to **determine** the whereabouts of the story's main character and his diamond.

We propose some suggestions to implement, develop, and adapt this task to specific levels and learner profiles:

- *Work with this task in the Spanish intermediate level or in subsequent levels.* The task can be guided to help students communicate more in Spanish. This guidance can take the form of a step by step introduction to the characters and their names, the places they find themselves in, and the actions that take place in the story where vocabulary related to places (e.g., *laboratorio*) and objects (e.g., *carbón, diamante*) is provided and practiced before the task. After the students have completed the task and reached their conclusions, a guided in-class discussion of the events in the video whereby students justify their conclusions will allow them to orally express them further, listen to other classmates, and make adjustments to their perceptions of the task if needed.
- *During planning time, focus on the review of vocabulary and pronunciation of complex words* by reading a news report about a robbery that the students need to then summarize orally in their own words working collaboratively. This pre-task activity can foster students' improvement of fluency of speech production.
- *Create handouts (one for Student A and one for Student B) with some guiding questions related to the information in the videos*. Written questions such as "What are the professions of the characters in your video? Why is the diamond important in the story? What happened to César Cabral and where is he now?" will allow students some time to remember the structure and actions in the videos, reflect on the events of the story, and have something ready to present to their partners when they start the task. Preparation and reflection time will give way to a smoother description of the task by both parts.
- *Take into account students' orientation toward the task, and make instructions clear*. For example, it is possible that some students see the task as a game they have played in the past and, therefore, may not communicate much. Other students might see the task as one that mandates the use of certain linguistic structures and may likewise limit communicative

interaction. Therefore, providing clear instructions on task outcomes will enable enhanced interaction during task completion.

- *Manipulate the task to target the pronunciation of particular words that contribute to the aim of the task*, such as multisyllabic words. You may choose to add those words in the video or, if subtitles are included, highlight them in a different color. In this way, the task can be altered to be a focused task where attention to multisyllabic words, combined with the task repetition element, targets improved fluency. Additionally, consider recording new voice-overs to manipulate speech rate, pauses, prominence, stress, thought groups, etc.
- *Guide a follow-up discussion on regional variation, in this case on Peninsular and Argentinian Spanish, to encourage development of sociolinguistic competence*. The conversation can be introduced by asking what differences the students noticed in the pronunciation of words in the two videos. The words can be written on the board to provide specific examples for a brief explanation by the teacher about the pronunciation and intonation variation in both countries.

CONCLUSION

Tasks are believed to encourage attention to and meaningful practice of the target language for the purpose of facilitating its acquisition. Our task facilitates an environment for learning in several ways. For example, it requires information exchange given that learners cannot complete the task unless they exchange information. This information exchange, as well as the negotiation of the meaning that is driven by consensus reaching or collaboration (i.e., when learners have to come to an agreement on a decision; see Berwick, 1993) prompts input, interaction, and output, all key ingredients for language acquisition (Gass & Mackey, 2006).

Tasks further represent a promising avenue for meaningful practice that may facilitate pronunciation in L2s. These kinds of language-based interactions are not widely included in textbooks (to the best of our knowledge), and pronunciation-based activities need to be emphasized in language classrooms. In this teaching tip, we described a task designed for intermediate-level learners of Spanish to target fluency development. Considering the information gap, our task enables students to communicate the main events of a story that, in turn, generates extended speech production; this extended speech production is a precursor for the development of spoken fluency. By means of task repetition, our unfocused task is predicted to promote fluency, which, in turn, may promote global comprehensibility. A convenient feature of our task (and tasks in general) is that it can be manipulated in order to address different languages, language level, and area of focus, to name but a few.

ACKNOWLEDGMENTS

The authors are grateful to Carly Henderson, who offered helpful feedback on an earlier version of this manuscript. We also thank the reviewers for their helpful comments. All remaining errors are our own.

ABOUT THE AUTHORS

Avizia Long (PhD, Indiana University) is Assistant Professor of Spanish at San José State University. Her research interests include second language variation, second language phonetics and phonology, and pronunciation in task-based language learning and teaching. She is co-author of *Sociolinguistics and Second Language Acquisition: Learning to Use Language in Context* (Routledge, 2014).

Contact information:
Department of World Languages and Literatures
One Washington Square, CL 421
San José, CA, 95192
Email: avizia.long@sjsu.edu

Lorena Alarcon received a M.A. in Applied Linguistics from Texas Tech University. Her research interests include technology to teach second languages, pronunciation in task-based language learning and teaching, second language variation, second language testing, and second language evaluation.

Contact information:
Department of Classical and Modern Languages and Literatures
Lubbock, TX, 79410
Email: lorena-fabiana.alarcon@ttu.edu

Sergio Ruiz-Pérez is a Graduate Part Time Instructor and second year PhD student of Spanish Linguistics at Texas Tech University. He holds a M.A. in Foreign Languages from West Virginia University. He completed his undergraduate studies in Spain, in English studies, translation and interpreting. His research interests include second language acquisition and instruction, task-based language learning and teaching, and translingual practices.

Contact information:
Department of Classical and Modern Languages and Literatures
Lubbock, TX, 79410
Email: sergio.ruiz-perez@ttu.edu

REFERENCES

Berwick, R. (1993). Towards an educational framework for teacher-led tasks in English as a foreign language. In G. Crookes & S. Gass (Eds.), *Task-based language teaching* (pp. 97-124). Clevedon, UK: Multilingual Matters.

Bygate, M. (1996). Effects of task repetition: appraising the developing language of learners. In J. Willis & D. Willis (Eds.), *Challenge and change in language teaching* (pp. 136-146). Oxford: Heinemann.

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford, UK: Oxford University Press.

Ellis, R. (2009). Task-based language teaching: Sorting out the misunderstandings. *International Journal of Applied Linguistics*, *19*(3), 221-246.

Gass, S. M., & Mackey, A. (2006). Input, interaction and output: An overview. *AILA Review, 19*(1), 3-17.

Gilabert, R. (2007). Effects of manipulating task complexity on self-repairs during L2 oral production. *IRAL-International Review of Applied Linguistics in Language Teaching*, *45*(3), 215-240.

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.

Jung, Y., Kim, Y., & Murphy, J. (2017). The role of task repetition in learning word-stress patterns through auditory priming tasks. *Studies in Second Language Acquisition, 39*(2), 319-346.

Kim, Y., & Tracy-Ventura, N. (2013). The role of task repetition in L2 performance development: What needs to be repeated during task-based interaction? *System*, *41*(3), 829-840.

Lambert, C., Kormos, J., & Minn, D. (2017). Task repetition and second language speech processing. *Studies in Second Language Acquisition*, *39*(1), 167-196.

Loewen, S., & Isbell, D. R. (2017). Pronunciation in face-to-face and audio-only synchronous computer-mediated learner interactions. *Studies in Second Language Acquisition*, *39*(2), 225-256.

Long, M. H. (1985). A role for instruction in second language acquisition: Task-based language teaching. In K. Hyltenstam & M. Pienemann (Eds.), *Modelling and assessing second language acquisition* (pp. 77-99). Clevedon, UK: Multilingual Matters.

McBride, K. (2015). Which features of Spanish learners' pronunciation most impact listener evaluations? *Hispania, 98*(1), 14-30.

McKinnon, S. (2017). TBLT instructional effects on tonal alignment and pitch range in L2 Spanish imperatives versus declaratives. *Studies in Second Language Acquisition, 39*(2), 287-317.

Parlak, Ö., & Ziegler, N. (2017). The impact of recasts on the development of primary stress in a synchronous computer-mediated environment. *Studies in Second Language Acquisition, 39*(2), 257-285.

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.

Skehan, P. (1998). Task-based instruction. *Annual Review of Applied Linguistics*, *18*, 268-286.

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.

Yule, G., Powers, M., & McDonald, D. (1992). The variable effects of some task-based learning procedures on communicative effectiveness. *Language Learning*, *42*, 249-277.

McGregor, A. (2019). Teaching Tip: Using selfies to improve problematic English consonants. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Ames, IA, September 2018 (pp. 393-400). Ames, IA: Iowa State University.

TEACHING TIP

USING SELFIES TO IMPROVE PROBLEMATIC ENGLISH CONSONANTS

Alison McGregor, Princeton University

The pronunciation of English consonants is often taught on an articulatory basis, focusing on static descriptions of the segments (sounds) and their distinctive features. A limitation in typical English consonant training, however, is that it may fail to incorporate any explicit instruction on the underlying mechanisms that trigger and coordinate the movements of the articulators. This teaching tip introduces lip aperture and protrusion related to producing and differentiating the commonly problematic English consonants /l/, /r/, /w/, /ʃ/, /ʒ/, /tʃ/ and /dʒ/. Based on the underlying positions and movements, consonant lip-rounding categories for these sounds are proposed to simplify the noticing and understanding of the subtle underlying production mechanism. For explicit pronunciation instruction, teachers can use the selfie classroom activity described in this teacher tip to systematically raise awareness, provide explicit instruction, create a metalanguage, and practice the mechanisms behind the production of these English consonants.

INTRODUCTION

Common challenges for learners from at least 14 different L1s include the struggle to accurately produce the following consonant sounds: /l/, /r/, /w/, /ʃ/, /ʒ/, /tʃ/, and /dʒ/ (Avery & Ehrlich, 1992). Pronunciation instruction of these consonants is often lacking information on the underlying mechanisms that enables articulation. According to articulatory phonology, "gestures are events that unfold during speech production and whose consequences can be observed in the movements of the speech articulators" (Browman & Goldstein, 1992, p. 23). A gesture can be thought of as the triggers or activities behind the positioning and motioning of the articulators (top/bottom lip, tongue, jaw). In this teaching tip, AG will be limited to the activities revolving around the lip involvement when producing the problematic English consonants /l/, /r/, /w/, /ʃ/, /ʒ/, /tʃ/, and /dʒ/ and will refer specifically to lip aperture (the actions of the upper lip, lower lip, and jaw) and lip protrusion (the degree to which the lips stick out).

An example of the cross-linguistic variation of underlying gestures can be found in a comparison of the Mandarin consonants written in Pinyin: *j*, *q*, and *x*, (e.g. /tɕ/, /tɕʰ/, /ɕ/) versus the English consonants j, ch, and sh (e.g. /dʒ/, /tʃ/, and /ʃ/). The visual cues for the Mandarin sounds, for example, are a spread open-smile position[i], while in contrast, the English consonants show more puckering and a slightly rounded lip position. These almost opposite gestures are due to the difference in aperture and protrusion—open/spread (not sticking out) in Mandarin versus more closed/slightly rounded (subtle puckering) in English—and demonstrate the difference in underlying mechanism used in sound production and coordination. Without explicit instruction, learners will not be aware of these "articulatory routines," which started as discrete gestures (babbling) in childhood and became gross gestures in adulthood used to differentiate and coordinate speech production. For a Mandarin speaker of English, these gesture differences not only present production inaccuracies, but also create variation in the visual cues for listeners

(Bikerman, 2014; McGurk & MacDonald, 1976) and contribute to the lack of coordinating movements necessary for making connected speech.

To help teachers guide learners, this teaching tip proposes the description and use of consonant lip-rounding categories to facilitate the learning and differentiation of the problematic English consonants /l/, /r/, /w/, /ʃ/, /ʒ/, /tʃ/, and /dʒ/. Note that lip rounding is typically a feature of English vowels (rounded or unrounded), but here it describes the outcome of articulator triggers to make the positioning and motioning of articulators in consonant production. Unlike distinctive features, gestures do not map onto segments or features but underlie the articulator coordination that make them (Browman & Goldstein, 1992). *Consonant lip rounding* refers to the underlying gestures of lip aperture and protrusion in the production of these English consonants.

As a pedagogical technique, a selfie activity is recommended to start the process of orienting learners to gestures. A selfie is a digital self-portrait shared on the Internet (Johnson, Maiullo, Trembley, Werner, & Woolsey, 2014). Due to their popularity, selfies are being used as a pedagogical tool to engage learners in a variety of ways, including the ice-breaker selfie (Johnson et al., 2014), the mathematical selfie (Jaqua, 2017), and the post-library-instruction selfie (Meehlhause, 2016). Merits of the selfie pedagogical tool include increasing student engagement, building classroom community, bringing authentic real-world material into the classroom, and providing an assessment tool. In the current teaching tip, the selfie is used for raising students' awareness of lip positions and movements and metalanguage (having a language to talk about those target actions). Using selfies for noticing positions and movements is a starting point that can be followed with highlighting the descriptive language for classroom instruction and feedback (metalanguage) about lip rounding can be established, paving the way for explicit instruction and feedback to occur throughout the class.

Teachers are most effective when they are able to guide learners in finding and coordinating underlying motor skills to make sounds, words, and ultimately fluent speech production (Catford & Pisoni, 1970). Below, I describe what teachers need to know about (a) common English consonant problems, (b) lip rounding for these problematic consonants, and (c) a selfie classroom activity to introduce and practice consonant lip rounding.

WHAT TEACHERS NEED TO KNOW

Common pronunciation challenges by L1s

Specific language groups have pronunciation challenges with the English consonants /l/, /r/, /w/, /ʃ/, /ʒ/, /tʃ/, and /dʒ/. Table 1 lists 14 languages with their respective common English consonant problems (Avery & Ehrlich, 1992).

Table 1

Common L1 consonant problems related to lip rounding

	Letters	L	R	W	SH		CH	J
	Sounds	/l/	/r/	/w/	/ʃ/	/ʒ/	/tʃ/	/dʒ/
LANGUAGE	Arabic		✓ (trill)				✓	✓
	Chinese	✓		✓/w/ vs. /r/	✓	✓	✓	✓
	Farsi		✓ (trill)	✓/w/ vs. /v/				
	French		✓	✓/w/ vs. /r/			✓	✓
	German			✓/w/ vs. /r/				✓
	Greek		✓			✓	✓	
	Hindi			✓/w/ vs. /v/				
	Italian		✓(trill)		✓	✓		✓
	Japanese		✓/r/ vs. /l/		✓/ʃ/ vs. /s/		✓/t/ vs. /tʃ/	
	Korean		✓/r/ vs. /l/		✓/ʃ/ vs. /s/			✓
	Polish	✓ /l/ vs. /w/	✓(trill)		✓/ʃ/ vs. /s/			
	Portuguese	✓ /l/ vs. /w/				✓ /ʒ/ vs. /dʒ/	✓	✓
	Spanish		✓(trill)		✓/ʃ/ vs. /s/			✓
	Vietnamese						✓	

Consonant lip rounding

In order to prepare learners to produce and differentiate problematic consonants, three types of lip rounding can be introduced. The titles—no lip rounding, slight lip rounding, and tight lip rounding (Edwards & Strattman, 1995)—describe the lip positions, which are the consequence of accurate underlying articulator mechanisms. In Table 2, the three types of lip rounding are displayed and described to show the fundamental differences in the production of the problematic consonants. It is important to note that these are static pictures but not static positions in speech. In other words, in the production of speech, lips will naturally move into and out of positioning to create target sounds and move on. Consonant lip rounding will naturally be influenced by surrounding vowels, for example, in the word "low" a no lip rounding /l/ would show lip rounding from the vowel /o/. It is because of these influences that gains in basic consonant lip-rounding accuracy will improve not only accuracy in the production of a target consonant, but also the development of connected speech (linking) since gestures enable the dexterity of articulators.

Table 2

Lip rounding categories and sample pictures

Type of Lip Rounding	No Lip Rounding	Slight Lip Rounding	Tight Lip Rounding
Gesture			
Description	Make the initial /l/ sound, as in "like," and you will notice the mouth opens slightly to allow room for the tongue to position itself. The exterior of the lips shows no rounding.	Make the "sh" sound or say "Sssh!" You will notice the corners of the mouth engage and pucker, although the top of the lip does not move. The exterior of the lips shows a slight lip-rounding position.	Make the /w/ sound, as in "what," and you will notice the top and bottom part of the lip engage strongly. The exterior of the lips shows a tight rounding position.
Differences	There is no engagement of exterior lip muscles, although the jaw drops ever so slightly.	The corners of the lips engage.	The top and bottom of the lips engage tightly.

THE SELFIE TEACHING TIP

Tip: Students will use cell phone cameras to take selfies of specific facial expressions and will then be asked to notice and describe the positions of the lips in the pictures.

Level: Appropriate for all levels

Description: This activity draws students' attention to the lip-rounding positions and motions that facilitate familiarity with both the introduction of lip rounding for English consonants and the language to talk about these position and motions.

Prerequisites: Students will need to be familiar with these terms: *smile*, "*Shhh!*" and "*What?!*"; *mouth*, *corners of mouth*, *lips*, *upper/bottom/top lip*, *spread lips*, *protruding (sticking out)*, *round*, *engage*, and *muscles*.

Materials: Cell-phone cameras in selfie mode; the list of facial expressions to make in each selfie (below); a list of adjectives to describe the lips (above)

In-class selfie activity instructions

1. Selfie time

Tell students to get out their cell phones, open the camera app, and set it to selfie mode. Instruct them to take three pictures doing the following: a) smiling, b) saying, “Shhh!” and c) saying “What?!,” stopping at the beginning of the word to capture making the /w/ sound.

2. Describe and discuss

After the selfies are taken, ask students to look at each picture and describe what they notice about their lips. To facilitate description and discussion, the following options are recommended, depending on the amount of available time and the best fit for the target group. Use think/pair/share: Have each individual jot down his or her thoughts, share them with a partner, and then discuss as an entire group. Alternatively, the instructor can model doing a think-aloud (to review adjectives) or students can discuss in small groups or describe and discuss as an entire group. The aim of this component of the activity is for learners to notice the differences in the lip positions and learn the descriptive language to talk about those positions and motions. Below are examples of types of descriptive points to highlight:

a) Smile Selfie: Lips are spread; mouth may be closed, teeth may or may not show; corners of the lips go in (into the cheeks), not out. [Note: In versus out is very important for the next step.]
b) “Shh!” Selfie: Muscles in the corner of the mouth are “on” (engaged/constricted); bottom lip and corners of lips stick out more than the top lip. [Note: Top lip should not stick out much in the next step.]
c) “What?!” Selfie: Lips are tightly rounded to make /w/; top and bottom lip are equally engaged. [Note: The top lip muscle is “on” in the /w/ selfie, as opposed to the top lip being more “off” in the “Shhh!” selfie.]

3. Teaching three categories of lip rounding for consonants

Now that students are aware of differences in lip rounding, choose the common problem consonants or target sound(s) and present explicit instruction on the lip motion and characteristics for each sound (see Table 2).

4. Practicing lip rounding

In general, it is recommended that instructors assess needs, including which sound(s) are problematic and why (which characteristics are causing the problem); raise awareness of the letter/sound/spelling and lip motions and provide explicit instruction. Next, structure the move from step-by-step in “finding the sound”—that is, finding the coordination—to putting the lips in the position and making the movement. After the students make it, focus on helping them make it effortlessly and automatically.

Trouble-shooting

1. **No lip rounding:** For learners whose lip position is accurate, ask them to relax the tongue to allow it to spread a bit more (this works particularly well for Japanese, Mandarin, and Korean speakers). Next, try to figure out where the tongue is hitting—behind the teeth, on the alveolar ridge, or above the alveolar ridge—and identify which part and how much of the tongue is making contact (see Raver-Lampman & Wilson, 2018).

2. **Slight lip rounding:** Two errors are common when learners attempt to make slight lip rounding. First, when they trigger their muscles, the corners of their lips will tend to go "in" as in the smile position. A good pre-exercise is to have students look in a mirror or camera and do an in/out motion–smile and then use the same muscles in the opposite direction so that the lips go out (something like a pout). Ask students to do this in a controlled manner in repetitions of five times in a row until it becomes easy. Next, ask them to just make the out motion. Again, repeat it five times or until it becomes controlled. A second error is that the learner will move the top lip too much or it will stick out like a duck bill. The instructor can ask the learners to put their index finger on the top lip to keep it from moving or use a finger or pen pressed gently against the center of the lips to hold down the top. This muscle isolation—moving the corners of the lips without rounding the top of the lip—can be quite challenging but underlies multiple consonants and promotes connected speech. If the top lip moves and constricts significantly, then the "sh" can sound like the /w/ sound. To differentiate, the muscle isolation (not moving the top lip) will need to be mastered.

3. **Tight lip rounding:** The most common error for this motion is that learners make the sound before getting into the position. The sound will then turn into "uuwh." To help learners, be sure to instruct them to get in the position before making the sound. You can practice this by making the position without making any sound. Then, in a slow-motion drill, say, "Get in position…. (watch for tight lip rounding) …okay, now make the sound." They also do not need to pull back on the lips to get out of this position, but just simply relax the lips. In rapid speech, however, the motion will be influenced by the upcoming sounds.

Sound-spelling correspondence

An additional barrier to pronunciation of problematic consonants is the lack of sound-spelling correspondence. In other words, English spelling often fails to transparently convey what the L2 learners' articulators need to do to make sounds and word-level pronunciation. As a follow-up activity, teaching sound-spelling correspondence for the problematic consonants is recommended. Table 3 highlights the sound-spelling correspondence (Grant, 2010) and shows example words with the respective lip-rounding type. Teaching learners to translate spelling into accurate sound information for lip positions will trigger the gestures behind the production of the sounds and reduce some of the frustrating mystery behind pronunciation of these English consonants.

Table 3

Lip rounding categories with sound-spelling correspondence

Lip Rounding	Letter	Sound	English Spelling Correspondence	Examples
1. No	l	/l/	l-ll-	**l**isten, a**ll**ow, ab**l**e, fu**ll**
2. Slight	sh	/ʃ/	-sh-ti-ci-ssi/ssu-si	Engli**sh**, mo**ti**on, so**ci**al, i**ssu**e, ma**ch**ine
		/ʒ/	-g-si/su-	Gara**g**e, A**si**a, mea**su**re
	ch	/tʃ/	-tu-	Fea**tu**re, si**tu**ation
	j	/dʒ/	-du-	Indivi**du**al, gra**du**ate
	r	/r/	-r*	Row**,** **r**eason, e**rr**o**r**, ea**r**
3. Tight	w	/w/	-w-wh	**wh**at, **wh**y, a**w**are, follo**w**
			qu	**qu**iet, **qu**estion, e**qu**ation

CONCLUDING REMARKS

This teaching tip integrates instruction of articulatory gestures related to lip aperture and protrusion specific to problematic English consonants with a practical classroom technique for pronunciation improvement. Although an articulatory approach to English consonant training offers explicit information, learners may need additional help using their articulators to orchestrate the production of problematic English consonants. Adding instruction related to lip aperture and protrusion gives learners the opportunity to better understand what to change in order to have more success in learning difficult consonants. Using selfies as a pedagogical tool offers a fun and practical technique to introduce positions and motions of underlying mechanisms without ultrasound or using an MRI of underlying language specific gestures. Teachers are encouraged to practice and experiment with the demonstration of consonant lip rounding, as well as to observe their learners' L1 gestures in consonant production.

ACKNOWLEDGEMENTS

I would like to acknowledge the collaboration between David Alan Stern and the late Howard Edwards for coming up with consonant lip rounding categories.

ABOUT THE AUTHOR

Alison McGregor is an instructor and testing coordinator for the English Language Program in the McGraw Center for Teaching and Learning at Princeton University. She teaches, tutors, and tests oral-communication skills for international graduate students and postdoctoral fellows. Her applied linguistics research focuses on the comprehensibility of oral English, factors in effective pronunciation instruction, and American English intonation.

Contact information:
Alison McGregor
Princeton University
87 Prospect
Princeton, NJ 08540
(609) 258-6928
Email: alison.mcgregor@princeton.edu

REFERENCES

Avery, P. & Ehrlich, S. (1992). *Teaching American English pronunciation.* Oxford, UK: Oxford University Press.

Browman, C., & Goldstein, L. (1992). Articulatory phonology: An overview. *Haskins Laboratories Status Report on Speech Research*, *111/112*, 23-42.

Bikerman. (2014, June 26). The McGurk effect [YouTube video]. Retrieved from https://www.youtube.com/watch?v=yJ81LLxfHY8&t=2s

Catford, J., & Pisoni, D. (1970). *Auditory vs. articulatory training in exotic sounds* (BR-9-7740). Ann Arbor, MI: Center for Research on Language and Language Behavior. Retrieved from ERIC database (ED042174).

Edwards, H., & Strattman, K. (1995). *Accent modification manual: Materials and activities.* Independence, KY: Cengage Learning.

Grant, L. (2010). *Well said: Pronunciation for clear communication* (3rd ed.). Boston, MA: Heinle Cengage Learning.

Johnson, S., Maiullo, S., Trembley, E., Werner, C., & Woolsey, D. (2014). The selfie as a pedagogical tool in a college classroom. *College Teaching*, *62*, 119-120.

Jaqua, K. (2017). Mathematical selfies: Students' real-world mathematics. *The Mathematics Teacher, 111*(1), 54-59.

Meehlhause, K. (2016). Two parts reflection, one part selfie. *Communication in Information Literacy, 10*(1), 14-22.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(23), 746-764.

Raver-Lampman, G. & Wilson, C. (2018). An acceptable alternative articulation to remediate mispronunciation of the English /l/ sound: Can production precede perception? *TESOL Journal 9*(1), 203-223.

[i] The difference also includes variation in tongue position, voicing and aspiration.

Reed, M. (2019). Teaching Tip: Listening skills instruction: Practical tips for processing aural input. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Ames, IA, September 2018 (pp. 401-412). Ames, IA: Iowa State University.

TEACHING TIP

LISTENING SKILLS INSTRUCTION: PRACTICAL TIPS FOR PROCESSING AURAL INPUT

Marnie Reed, Boston University

Two listening challenges faced by English L2 learners are (1) successfully identifying words in continuous speech and (2) understanding a speaker's intended meaning. Listening is a skill L2 learners report wanting to improve, yet teaching practices often fail to advance learner knowledge and control of listening processes. Instructors can benefit from empirically-supported recommendations to help learners parse continuous speech, and discern speaker intent. This Teaching Tip shares two 3-part strategies to facilitate processing utterance content and interpreting message meaning. The practical tips presented here are consistent with a return in the larger TESOL field to a true communicative approach, relying on authentic materials and real communicative contexts rather than mere mimicry of connected speech features or particular intonation contours.

INTRODUCTION

Skilled listening is an essential part of communication, one which facilitates the emergence of other skills. Yet current work in second language pronunciation pedagogy suggests that listening is the subset of 'pronunciation' that is still earning the dubious distinction of "neglected orphan" (Deng et al., 2009; Derwing & Munro, 2005; Gilbert, 2010) relative to other skills. This is striking in light of its proportional importance. As noted by Nunan (1998), "over 50% of the time that students spend functioning in a foreign language will be devoted to listening" (p. 1). Despite this prominence, listening has also been described as "the least understood and most overlooked of the four skills in the language classroom" (Nation & Newton, 2009, p. 37). According to Vandergrift and Goh (2012), listening is the skill over which learners feel the least control and for which they receive "the least systematic attention from teachers and instructional materials" (p. 4). Indeed, most listening skills textbooks tend to bypass listening instruction, and focus instead on note-taking, a skill which presupposes the ability to comprehend the listening input. Finally, listening is found to be the skill for which teachers have received the least training (Graham, 2017; Siegel, 2014).

Novice teachers–or even seasoned teachers newly tasked with teaching listening–might turn to their institution's curriculum guidelines for direction. In an academically-oriented intensive English program steeped in the communicative language teaching (CLT) approach, teachers may discover listening objectives that are not operationalized ("Learners will understand…" does not answer the question, *As measured how?*) and that parallel reading objectives. Such 'objectives' provide no indication of where to start or how to go about the task. A representative sample of Guidelines for high-intermediate to advanced-level learners in a typical CEA-accredited intensive English program illustrates the conflation of reading and listening skills objectives. Consider the requisite skill to meet this reading objective: *Understand the main ideas and significant details.* Next, consider the requisite skill to meet this listening objective: *Understand main points and the* *most significant details*. To meet the reading objective requires ability to decode orthographic print, and pre-supposes working with literate students. As anyone who has worked with pre-literate students from limited or interrupted formal education (SLIFE) populations has discovered, you cannot simply hand out a passage with instructions to read and answer the comprehension questions. Learners who cannot read cannot process orthographic input. A comparable insight regarding teaching listening seems to have eluded us; that is, in the absence of instruction on how to process aural input, you cannot direct students to listen to a passage and expect them to process or fully comprehend it.

To recap, listening is an essential but difficult skill. In CLT classes, the focus of listening instruction seems to be the end product of comprehension. That is, as captured by Mendelsohn (2006), "Much of what is traditionally mis-named *teaching* listening should in fact be called *testing* listening" (p. 75). Describing this as a text-oriented approach attributable to the influence of traditional reading pedagogy, Vandergrift and Goh (2012) noted in a review of listening instruction that, "Instead of teaching how to listen accurately, listening activities tested the accuracy of learners' comprehension" (p. 6). Fortunately, however, listening has been receiving priority in empirical studies (Vanderplank, 2013), with a concerted effort underway to translate research into practice. As advocated by Field (2008), newer pedagogical guidelines call for listening to be taught as a language skill in its own right.

Two challenges for L2 listeners have been identified (Goh, 2000), and these mirror acquisition challenges for children acquiring their first language(s): parsing a continuous speech stream, and understanding speaker intent. Consider examples for English L1: the oft-cited mishearing by children of a Christian hymn, "Gladly the cross I'd bear", as a song about a cross-eyed bear named Gladly, or the exchange reported by Berko Gleason and Ratner (2009) of the first-grader bragging that the third-grader on his school bus was impressed by his new back pack because the older boy had said, "Big deal". The following examples illustrate these two challenges for English L2 learners.

Listening Challenge 1: Parsing connected speech to understand utterance content

Think of this as the 'izzybizzy' [IzibIzi] phenomenon. None of the three words in isolation is likely to pose problems for even beginning-level English-learning students; yet in connected speech, these words do not sound like they look in citation form. As experienced by one of our newly-arrived master's students some years ago who had cleared customs at the airport, retrieved his baggage, and found the taxi stand, he was flummoxed when the man in front of him in line asked, "Do you mind if I smoke?" to which he wanted to reply–or at least look up the word, but had no idea how to spell it since, as he put it to us, "What's a [maindIfai]?" The result of inability to segment continuous speech is that learners fail to recognize known words in rapid discourse.

What accounts for this may be the decoding strategies learners use to understand utterance content. Field (2008) notes that learners miss more function words than content words, perhaps attributable to adhering to what they are taught in test-prep classes: Pay attention to the content words, the little words aren't important. They may also unconsciously be applying L1 segmentation strategies to locate word boundaries. However, cross-linguistic differences in world-boundary acoustic cues adversely impact L2 segmentation discourse (Altenberg, 2005, Carroll, 2004). Even advanced learners have been shown to transfer L1 phonotactics even when the L1 phonotactics are not helpful (Al-jasser, 2008; Weber & Cultler, 2006). According to Broersma and Cutler (2008), learners also substitute known words for unrecognized words and have difficulty suppressing wrong choices. To illustrate what this looks like in practice, examine this sample dictation from a pronunciation elective class:

Challenge 1, Illustrative sample: Dictation task

Teller all meter at the bank.

The advanced-level students who had (mis)transcribed this utterance vigorously insisted that this is what I had dictated, all the while acknowledging that it made no sense.

Listening Challenge 2: Interpreting English intonation to understand speaker intent

The source of this challenge has been captured by Tomlinson and Bott (2013): "Often what a speaker intends to say is not directly retrievable from a linguistic form; rather listeners must infer it" (p. 3569).

To illustrate how this plays out, consider a common classroom exchange, reported in Reed and Michaud (2015):

Challenge 2, Illustrative sample: Student – teacher exchange

Student: "Teacher, can I turn in my assignment late?"
Teacher: "You *can*."
Student: "Okay, thanks!"

The teacher's words are affirmative, yet the message is negative. L2 listeners who miss the point of an utterance may be relying solely on the words, unaware of the signaling function of intonation. Wichmann (2005) accounts for the seeming contradiction of affirmative words conveying a negative message by ascribing to intonation "the power to reinforce, mitigate, or even undermine the words spoken" (p. 229). The native-speaker listeners are sensitive to what Wells (2006) refers to as the implicational fall-rise pitch contour whereby "a speaker implies something without necessarily putting it into words […] Something is left unsaid–perhaps some kind of reservation or implication" (p. 27). The non-native listeners, on the other hand, may be unconsciously applying L1 pragmatic interpretation procedures to comprehend conversation implicatures or to go beyond the literal meaning of an utterance (Cutler, 2001; Garcia, 2004). When listeners fail to attend to prosodic cues, the result is that they fail to grasp the message despite understanding the words.

Listening skills instruction is called for. In listing key pedagogical principles, Graham, Santos, and Francis-Brophy (2014) report zero or slow learner progress without listening instruction, but effective development with it. In light of research that suggests that less-skilled listeners rely on bottom-up processing (Tsui & Fullilove, 1998), Graham et al. recommend strategies for building, verifying and monitoring bottom up processing in order to augment popular top-down pre-listening Strategies that relate prior knowledge to listening passages. And in keeping with Goh (2002), they recommend metacognitive strategy awareness and use.

Instructors, too, need strategies to help break down the components of listening in order to teach learners how to actually process spoken input. Metacognitive strategy-based training in connected speech has been found to increase learner awareness and skills necessary to aid word segmentation, and training in contrastive stress and intonation has been found to facilitate understanding speakers' intended meaning as well as the message. To address these two barriers to listening comprehension, Teaching Tips are offered to facilitate development of segmentation skills that allow recognition of known words in connected speech, and to promote awareness of the discourse and pragmatic functions of intonation that allow inferring what is meant by what is said. So where do we start? Two 3-part strategies are shared below to facilitate processing utterance content and interpreting message meaning. The practical tips presented here are consistent with a return in the larger TESOL field to a true communicative approach, relying on authentic materials and real communicative contexts rather than mere mimicry of connected speech features or particular intonation contours.

Understanding what was said

English doesn't sound the way it looks. In continuous discourse, words do not retain their dictionary citation forms. Unlike the written page, there's no white space between the words. Decoding Connected Speech requires knowledge of Connected Speech Processes (CSPs). Teaching Tip: introduce CSPs for in-class practice and incorporate them into your lessons. The figure below provides a one-page see-at-a-glance select list of CSPs with accompanying examples.

Connected Speech Processes: ***English doesn't sound the way it looks.***
(important for listening comprehension; not necessary for speaking)

ɪzzybɪzzy ***How many words? Where are the word boundaries?***
Words don't sound like their dictionary entries.

Sounds are linked

Consonants to consonants	(same place of articulation: *sit down, last time, some more*)
Consonants to consonants	(different place of articulation: *last page, social media*)
Consonants to vowels	(*take on, talk about*)
Vowels to vowels	(*key issue; go around*)

Sounds are deleted

/h/ in *he, her, his,* and *him*	***Except:*** Sentence-initial: *He's late.*
E.g. *Is he busy?* → *Is ~~he~~ busy?*	Stressed: *I don't mean him; I mean her.*
Sounds like: *[ɪzzybɪzzy]?*	Pronouns: his/ hers: *That's not his; it's hers.*

Sounds are reduced

can → kn	I can do it. → *I kn do it.*
and → n	Law and Order → *Law 'n Order*; wait and see → *wait 'n see*
an → n	US units (mph) miles per hour → *miles an hour* → *miles 'n hour*
of → ə	a lot of time → *a lotta time*
or → r	right or wrong → *right 'r wrong*

Sounds are altered

you → yə	See you later → *See ya later.*
After /d/ → jə	Would you? → *Would ja?* Could you? → *Could ja?* Did you? → *Did ja?*
After /t/ → chə	Can't you? → *Can'tcha?* Won't you? → *Won'tcha?*
to → tə	have to → *hafta*
/t/ → D	letter → leDer, better → beDer, water → waDer
you're/your → yer	You're right. → *yer right.* It's on your right. → *It's on yer right.*
After /d/ → jer	You did your homework? → *You di djer homework?*
After /t/ → cher	Put your hat on. → *Pu tcher hat on.*
got to → *gotta*	
want to → *wanna*	
going to → *gonna*	

Words're contracted

(Negative) not → n't*: isn't, aren't, doesn't, don't, won't, can't, shouldn't,* etc.
(Auxiliaries): I am→ *I'm,* I will→ *I'll,* I have→ *I've,* I would/had→ *I'd,* I would have→ *I'd've*
you are → *you're,* you have → *you've,* you will → *you'll,* you would/had → *you'd*
(s)he/it/they→ he's, she's, he'll, she'll, it'll, he'd, she'd, it'd, they're, they'll, they've, they'd
(Modals): could have→ *could 've*, would have→ *would 've*, should have→ *should 've, etc.*
(Existential Pronouns): there is a → *there's a*; there are → *there're*
(Proper Nouns): Jane will …→ *Jane'll;* Bob will → *Bob'll . . .*
(Common Nouns): the judge will . . . → *the judge'll . . .*

Figure 1. Frequently occurring connected speech processes.

A role for auditory feedback

In English-as-a-Second Language settings, that is, when students are learning in an English-speaking country, or when students are accessing YouTube or other available resources, they are exposed to the features whereby words are linked and contracted, sounds are reduced, deleted, and altered. To the extent that input exhibiting these CSPs is not understood in exposure-rich settings or under conventional listening instruction, it must be the case that external input is not a sufficient condition for accurate perception. As suggested by Casserly and Pisoni (2010), the alternative to a focus on perception is shaping the speaker's own speech production to activate robust auditory feedback. A number of studies that explore the conventional precedence of perception over production have demonstrated that production skills can exceed perception abilities (Sheldon & Strange, 1982). While acknowledging that Flege's (1995) Speech Learning Model accounts for acquisition of the majority of second language sounds, Linebaugh and Roche (2013) found that production training of problematic second language sounds improved perception, while additional listening exposure did not. Successfully extending their research to additional (perceptually assimilated) sounds, Linebaugh and Roche (2015) concluded, "We find compelling evidence that any model of second language phonological acquisition must accommodate the fact that production can inform perception" (p. A-9). Put simply, when learners' own speech production converges with the target pronunciation, an auditory feedback loop is created whereby "Speaking helps listening" (Reed & Michaud, 2010). Though not yet empirically investigated, it seems plausible that this extends to connected speech. Therefore, inform students that in-class practice producing these CSPs will facilitate out-of-class listening comprehension. Make clear that your students are not required to adopt these CSPs in their own out-of-class speech, nor will they receive error correction for not producing these CSPs in their spontaneous in-class speech.

Armed with knowledge of English CSPs, students are now better equipped to use the two 3-part Listening Strategies described below in order to segment continuous speech and understand the content of what was said.

Listening Strategy 1: Use three kinds of information to process aural input

1. *Use context information–what you already know about the topic of conversation: background knowledge, world knowledge, content knowledge.*

2. *Use language information–what you know about how the English language works: the grammar, the vocabulary, and the sound system.*

3. *Use acoustic information–the sounds that you actually hear someone saying.*

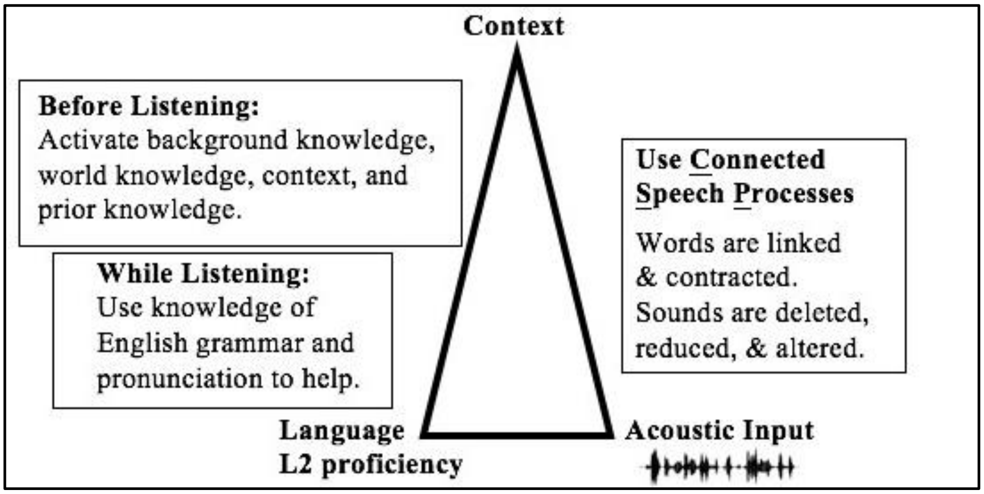


Figure 2. Strategy to facilitate processing utterance content.

Instructional debrief: Supplement strategy instruction with decoding practice

A common classroom activity-the dictation-can be a frustrating experience for students and a very humbling experience for teachers. This is particularly true with students who seem quite fluent when *they* speak, and who generally nod with seeming comprehension when *you* speak. Their transcriptions reveal the listening deficiency. Likewise, students' pleas for repeated playing of authentic material, such as snippets from podcasts or TED Talks, reveal their need for effective strategy implementation for efficient listening. The figure below offers three practical steps to implement the listening strategy.

TEACHING TIP: USE THREE STEPS TO PROCESS CONNECTED SPEECH
Step 1: What did you hear? Write down what it sounded like.
Step 2: Does it make sense? Reread what you wrote.
Step 3: What was really said? Use the three kinds of information to decode what you heard.

Figure 3. Three steps to decode aural input.

To illustrate the listening strategy in action, consider how to debrief the incorrect transcription below.

Dictation Example: Tell her I'll meet her…

Use 3 Steps to make sense of what you hear

Step 1: **What did you hear?** Repeat/write down.
Sounds like: *Teller all meter*.
Step 2: **Does it make sense?** Think/reread.
No, but this is what the spoken sentence sounds like.
Step 3: **What was really said?** Reconstruct/ Use 3 Kinds of Information

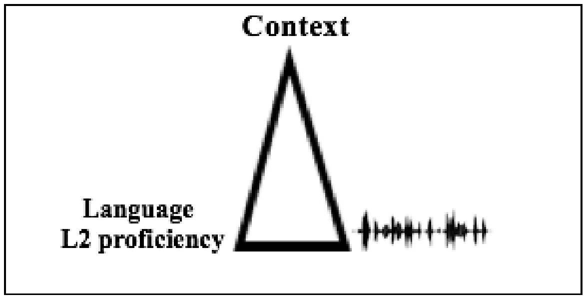


Figure 4. Three kinds of information to process aural input.

1. ***Top Down Processing***. There's no context.
   Background information cannot be activated.
2. Use **language information**.
   Every English sentence needs a verb. Possible verbs: 'tell' and 'meet'. Every verb needs a subject. In a command, the unspoken subject is 'you'. Most English sentences pattern **S**ubject–**V**erb–**O**bject; this one has 2 clauses. *(Will you please) tell her (that) I'll meet her.*
3. ***Bottom Up Processing*. Use acoustic (Sound) information.**
   Sounds are Deleted; /h/ ⇒ /Ø/ in *~~h~~e, ~~h~~er, ~~h~~is, ~~h~~im* except: when it's the first word in a sentence or clause; when it's stressed for emphasis; when *his* functions as a possessive pronoun.
   Words are contracted: *I will ⇒ I'll*
   Words are linked: *tell ~~h~~er ⇒ tell~~h~~er; meet her ⇒ meet~~h~~er (Will you please) tell ~~h~~er (that) I'll meet~~h~~er.*
   Tell ~~h~~er I'll meet ~~h~~er.

Understanding what was meant by what was said

The source of Listening Challenge 2 is perhaps best captured by Paunović and Savić (2008):

> "Students often do not have a clear idea of why exactly 'the melody of speech' should be important for communication, and therefore seem to lack the motivation to master it, while teachers do not seem to be theoretically or practically well-equipped to explain and illustrate its significance" (pp. 72-73).

As Gilbert (2014) observed, learners "will rarely tell the teacher they feel silly speaking this way, and the result will be that they may walk out of the class without having accepted the system at all. Or they may think intonation is simply decorative" (p. 125).

Listening Strategy 2: Use three kinds of information to decode speaker intent

Part I: Neutral, Unmarked Stress & Intonation*

Lexical: stressed syllable is longer, louder, higher (in pitch)
Phrasal: stressed content words; unstressed function words
Sentence: wide English neutral (unmarked) pitch range
Discourse: wide English neutral (unmarked) pitch range

Part II Marked Stress & Intonation

Sometimes, understanding what was said in English feels like trying to break a code.
You can use three kinds of information to help you decode a message.

1. ***Detect the Signal:*** *Marked Pitch Range (extra pitch).*

***Same*/Different**: Listen & Respond: Are the two sentences The Same or Different?

   a. She's not a *teacher.*
   b. *She's* not a teacher.

2. ***Locate the Signal***: *Exaggerated content or function word(s).*
   a. *The word 'teacher' had extra pitch.*
   b. The word "She's' had extra pitch.
3. ***Interpret the Signal***: *Attribute speaker intent.*
   a. She's not a *teacher*: She's an engineer.
   b. **She's** not a teacher: He's a teacher.

DISCOURSE & PRAGMATIC FUNCTIONS OF INTONATION

1: Emphatic Stress
Same meaning, just emphasized: $X \Rightarrow X^n$
2: Contrastive / Corrective Stress
±stated, Different Meaning: not X, Y
3: Implicational Stress
Unstated, not retrievable form the utterance alone: X + Y

Practice: Listen to the following two sentences; Answer the question; Explain your answer.

1. The teacher didn't' grade the papers.

Question: Have the papers been graded?

No.
Neutral stress & intonation.

2. The *teacher* didn't grade the papers.

Question: Have the papers been graded?

Yes, but not by the teacher.
The word 'teacher' had extra stress & intonation.

Figure 5. Strategy to facilitate interpreting message meaning.

As Levis (1999) cautioned, the historic textbook treatment of intonation is to overemphasize its role in signaling grammatical relations or its role in conveying speakers' attitudes and emotions. Instead, as Allen (1971) advocated, we should provide instruction that "teaches the student to think in terms of the speaker's intention in any given speech situation" (p. 73). To achieve this at the skill level, encourage students to practice *producing* marked stress and intonation in order to be able to *hear* marked stress and intonation. To facilitate this at the level of metalinguistic awareness, encourage students to articulate the communicative and pragmatic functions of intonation. Use metacognitive strategy instruction to introduce the three steps to process speaker intent: detect the aural signal (marked pitch range), locate the signal (exaggerated content or function word), and interpret the signal (emphatic stress, contrastive or corrective stress, or implicational stress). Finally, take advantage of tech tools to raise awareness and motivate practice.

SUMMARY

This Teaching Tip addresses two challenges to processing aural input. It advocates teaching connected speech processes to improve ability to segment continuous speech. It offers a metacognitive three-part strategy to process utterance content to understand what was said, and a metacognitive three-part strategy to process message meaning and interpret speaker intent.

ABOUT THE AUTHOR

Marnie Reed is a professor of education at Boston University. Her research focuses on second language phonology. Her current areas of interest include the role of metacognition in cross-linguistic awareness of the pragmatic functions of English prosody.

REFERENCES

Al-jasser, F. (2008). The effect of teaching English phonotactics on the lexical segmentation of English as a foreign language. *System, 36*(1), 94-106.

Allen, V. (1971). Teaching intonation, from theory to practice. *TESOL Quarterly, 5*, 73-81.

Altenberg, E. P. (2005). The judgement, perception, and production of consonant clusters in a second language. *International Review of Applied Linguistics, 43,* 53-80.

Berko Gleason, J., & Ratner, N. B. (2009). *The development of language*. (7th ed.). Boston, MA: Pearson Allyn & Bacon.

Broersma, M., & Cutler, A. (2008). Phantom word activation in L2. *System, 36*, 22-34. Carroll, S. E. (2004). Segmentation: Learning how to 'hear words' in the L2 speech stream. *Transactions of the Philological Society, 102*(2), 227-254.

Casserly, E. D., & Pisoni, D. B. (2010). Speech perception and production. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(5), 629-647.

Cutler, A. (2001). Listening to a second language through the ears of a first. *Interpreting, 5*, 1- 23.

Deng, J., Holtby, A., Howden-Weaver, L., Nessim, L., Nicholas, B., Nickle, K., Panneboek, C, Stephan, S., & Sun, M. (2009). *English pronunciation research: The neglected orphan of second language acquisition studies?* (Working Paper WPO5-9). Alberta, Canada: University of Alberta.Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly, 39,* 379-397.

Field, J. (2008). *Listening in the language classroom*. Cambridge, UK: Cambridge University Press.

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). Timonium, MD: York Press.

Garcia, P. (2004). Pragmatic comprehension of high- and low-level language learners. *TESL-EJ, 8*(2), 1-15.

Gilbert, J. (2010). Pronunciation as orphan: What can be done? *Speak Out!, 43,* 3–7.

Gilbert, J. (2014). Myth 4: Intonation is hard to teach. In L. Grant (Ed.), *Pronunciation myths: Applying second language research to classroom teaching* (pp. 107-136). Ann Arbor, MI: University of Michigan Press.

Goh, C. (2000). A cognitive perspective on language learners' listening comprehension problems. *System, 28*(1), 55-75.

Graham, S. (2017). Research into practice: Listening strategies in an instructed classroom setting. *Language Teaching, 50*(1), 107-119.

Graham, S., Santos, D., & Francis-Brophy, E. (2014). Teacher beliefs about listening in a foreign language. *Teaching and Teacher Education, 40*(1), 44-60.

Levis, J. (1999). Intonation in theory and practice, revisited. *TESOL Quarterly, 33*, 37-63.

Linebaugh, G., & Roche, T. (2013). Learning to hear by learning to speak: The effect of articulatory training of Arab learners' English phonemic discrimination. *Australian Review of Applied Linguistics, 36*(2), 146-159.

Linebaugh, G., & Roche, T. (2015). Evidence that L2 production training can enhance perception. *Journal of Academic Language & Learning, 9*(1), A1-A17.

Mendelsohn, D. (2006). Learning how to listen using learning strategies. In P. Gordon (Ed.), *Current trends in the development and teaching of the four language skills* (pp. 75-89). Berlin: Mouton de Gruyter.

Nation, P., & Newton, J. (2009). *Teaching ESL/EFL Listening and Speaking*. New York: Routledge.

Nunan, D. (1998). Approaches to teaching listening in the language classroom. *Proceedings of the 1997 Korea TESOL Conference* (pp. 1-10). Taejon, Korea: Korea Teachers of English to Speakers of Other Languages.

Paunović, T., & Savić, M. (2008). Discourse intonation—Making it work. In S. Komar & U. Mozetić (Eds.), *As you write it: Issues in literature, language, and translation in the context of Europe in the 21st century, 5*(1-2), 57-75.

Reed, M., & Michaud, C. (2010). *Goal-driven lesson planning for English as a second language teachers*. Ann Arbor, MI: University of Michigan Press.

Reed, M., & Michaud, C. (2015). Intonation in research and practice: The importance of metacognition. In M. Reed & J. Levis (Eds.), *The handbook of English pronunciation* (pp. 454-470). West Sussex, UK: John Wiley & Sons, Inc.

Sheldon, A., & Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics, 3*(3), 243-261.

Siegel, J. (2014). Exploring L2 listening instruction: Examination of practice. *English Language Teaching Journal, 68* (1), 22-30.

Tomlinson, J. M., & Bott, L. (2013). How intonation constrains pragmatic inferences. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the Annual Meeting of the Cognitive Science Society, 35*, (pp. 3569-3574).

Tsui, A., & Fullilove, J. (1998). Bottom-up or top-down processing as a discriminator of L2 listening performance. *Applied Linguistics, 19*(4), 432-451.

Vandergrift, L., & Goh, C. (2012). *Teaching and learning second language listening: Metacognition in action*. New York: Routledge.

Vanderplank, R. (2013). Listening and understanding. In P. Driscoll, E. Macaro, & A. Swarbrick (Eds.), *Debates in modern languages education* (pp. 53-65). Abingdon, UK: Routledge.

Weber, A. & Cutler, A. (2006). First-language phonotactics in second-language listening. *Journal of the Acoustical Society of America, 119*(1). 597-607.

Wells, J. C. (2006). *English intonation: An introduction*. Cambridge, UK: Cambridge University Press.Wichmann, A. (2005). The role of intonation in the expression of attitudinal meaning. *English Language and Linguistics, 9*(2), 229-253.

Richards, M., & Cotos, E. (2019). Teaching Tip: Segmental accuracy: A recommended training sequence for moving learners from accurate perception to accurate and automatic production in the stream of speech. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Ames, IA, September 2018 (pp. 413-422). Ames, IA: Iowa State University.

TEACHING TIP

SEGMENTAL ACCURACY: A RECOMMENDED TRAINING SEQUENCE FOR MOVING LEARNERS FROM ACCURATE PERCEPTION TO ACCURATE AND AUTOMATIC PRODUCTION IN THE STREAM OF SPEECH

Monica Richards, Iowa State University
Elena Cotos, Iowa State University

A key reason L2 learners struggle to pronounce new segmentals is because their L1 has trained them to hear L2 phonemes as allophonic. When learners cannot accurately *hear* a word's phonemic structure, they are able to self-assess their L2 pronunciation only by comparing their conscious knowledge of how the word should be pronounced with the physiological "feel" of their vocal organs. But is it possible to perform this task consciously on a regular basis? After all, L2 speakers must simultaneously engage in several additional, higher-level cognitive processes also harder in L2 than L1, for example, comprehending what others are saying, drawing connections between what is said to what is already known, planning what to say next and figuring out how best to say it. L2 learners must therefore develop the ability to self-assess *sub*consciously whether the phones they pronounce are categorized by the L2 as the phonemes they intend. Not only that, but their physical production of accurate L2 phoneme distinctions must become *habitual*. This paper therefore introduces a recommended training sequence for moving learners from accurate perception to accurate and automatic production of challenging L2 segmentals in the stream of speech.

INTRODUCTION

L2 segmental training frequently focuses on the mouth. However, pronunciation is an automatic skill including receptive as well as productive components. Just as the mind originally trained the mouth to produce reliably accurate segmental pronunciations in the L1, the mind is vitally important for re-training the mouth to produce reliably accurate segmental pronunciations in an L2. We therefore overview three key realities underpinning maximally effective segmental pronunciation pedagogy: *1) perception and production are linked*, *2) segmental pronunciation is a physical activity* and *3) segmental pronunciation is a habit*. On this foundation, we present a recommended training sequence grounded in these three key realities in order to maximize L2 learners' success in developing and automatizing new, more accurate pronunciation of segmentals.

UNDERSTANDING THE PERCEPTION/PRODUCTION LINK

The ear and mind can potentially learn to distinguish hundreds of sounds (MRT [magnetic resonance tomography] YouTube video). However, babies acquiring their L1 are exposed to the smaller subset of sounds found in that L1, and their minds begin developing sound categories accordingly (usually less than a hundred – Moran, McCloy, & Wright, 2014). Each phoneme, or mental label for a sound category, may include a range of phones (i.e., distinct physical pronunciations). For example, the English /l/ phoneme is not only produced as light/clear [l] but also syllable-finally as dark [ɫ]). In part, this is because a sound's environment, its surrounding sounds, may facilitate slightly varying realizations. Additionally, people's vocal tracts differ slightly in size, shape, etc., also producing minute variation in a given phoneme's pronunciation. The mind learns to disregard this "noise" in the input to a large extent. Nevertheless, unfortunately for L2 learners, sounds that pertain to the same phoneme label are sometimes not phonetically close at all. For example, in General American English, the phoneme /t/ has 5 distinct allophones (phones categorized as belonging to the same phoneme even though they are in fact articulated differently). Some allophones of General American /t/ are quite different from the prototypical voiceless alveolar stop we commonly think of as /t/, as demonstrated in Figure 1.

In spite of how some L1-defined allophones are quite different from each other phonetically, L1 child learners nevertheless develop mental sound categories enabling their minds to mirror for any allophone they hear their L1's phoneme categorization as efficiently as possible. This "tuning" of the mind to notice only L1-relevant aspects of any spoken language input maximizes the processing efficiency of L1 speech perception, but it complicates learning to categorize L1 allophones as phonemic in the L2, i.e., as phonetically different sounds belonging to different L2 phonemes. As a result, learners frequently hear L2 phonemes as allophonic (e.g., L1 Japanese speakers characteristically struggle to hear the difference between the English /ɹ/ and /l/ phonemes and L1 Korean speakers the difference between English /p/ and /b/ because their respective languages categorize these sound pairs as allophones) (Best & Tyler, 2007; Broersma & Cutler, 2008; Flege, 1995; Jenkins, 2000; Qian, Chukharev-Hudalainen, & Levis, 2018; Richards, 2012).

General American allophones of /t/	**Environment**	**Example**
aspirated [tʰ]	the beginning of words and stressed syllables	"time", "return"
flap [ɾ]	usually between a vowel (± /ɹ/) and an unstressed vowel or syllabic [l̩], [m̩], or [ɹ̩] – also, between any vowel (± /ɹ/) and a word-initial vowel (Vaux, 2000)	"water", "liberty", "bottom", "turtle", "thought it over", "report it immediately"
nasalized flap [ɾ̃]	substituting for "nt"	"winter", "front of. . ." *(i.e., the same phone used to pronounce /n/ in "winner" and "tunnel")*
glottal stop [ʔ]	between a vowel (± /l-ɹ/) and syllabic [n̩]	"button", "bought an. . . .", "carton", "Hilton" *(cf. the medial glottal stop in "Uh-oh")*
unaspirated [t]	everywhere else	"wait", "doctor", "multiply", "distance", "stop"

Figure 1. The five General American allophones of /t/.

It is important teachers realize that when learners fail to recognize phonetically different L2 allophones as belonging to the same phoneme, it is not only their L2 *listening*, but also their L2 *pronunciation* that is affected. After all, how much impact can a teacher's occasional pronunciation correction be expected to make if learners' continued misclassification of L2 phonemes as allophones prevents them from recognizing the *vast majority* of their mispronunciations of a particular problem segmental? Indeed, when learners cannot hear a word's phonemic structure accurately, they can self-assess their L2 pronunciation only by comparing their conscious knowledge of how the word *should* be pronounced with the physiological "feel" of their vocal organs. The impracticability of this for the stream of speech should be evident in light of how the mind must simultaneously accomplish several higher-level cognitive processes also much harder in the L2 than L1, namely:

1) Comprehending what others are saying;
2) Identifying connections between what others are saying and what one already knows;
3) Figuring out what ideas one wants to say next and how to express them (in terms of information structure, politeness, grammar, etc.).

It is therefore vitally important that teachers prioritize developing learners' ability to passively hear whether or not the L2 categorizes their various phone pronunciations as distinct phonemes.

One final factor impacting L2 sound perception and L2 segmental production is orthography. Many L2 learners have stronger reading than listening skills, resulting in their L2 sound perception being impacted by spelling. That is, the pronunciation L2 listeners sometimes perceive is that which the spelling of a word leads them to expect (Detey & Nespoulous, 2008; Erdener & Burnham, 2005). Understandably, they may also speak this spelling-derived pronunciation (Young-Scholten & Archibald, 2000). It is therefore vital teachers 1) recognize which words learners misperceive and mispronounce due to orthographic ambiguity and 2) facilitate learners' perceptive and productive acquisition of these words' standard spoken form.

Nevertheless, maximally effective segmental pronunciation pedagogy is grounded not only in teachers' understanding of the perception/production link, but also in their recognition that segmental pronunciation *is* a physical activity.

UNDERSTANDING SEGMENTAL PRONUNCIATION AS A PHYSICAL ACTIVITY

Many L2 speakers exhibit near-native proficiency in L2 skills other than pronunciation (Scovel, 1969). However, advanced L2 speakers frequently struggle to acquire standard L2 pronunciation in part because segmental pronunciation has an unmistakably physical component.

That is, even if L2 speakers successfully learn to hear new L2 phonemes as phonemes (versus allophones), they often experience great difficulty learning *how* to move their vocal organs differently than how they have always moved them before. In part, this is because L2 learners are accustomed to regularly adjusting their understanding of new vocabulary and even grammar, but many have had very little pressure to adjust their pronunciation since first learning to speak their L1 as a child. Teachers' efforts to develop students' conscious understanding – or metaphonological awareness – of *how* a target segmental is physically articulated can therefore be helpful (Ingels, 2010). Nevertheless, teachers must not only provide L2 learners metaphonological information about *how* a sound is physically articulated; they must also help learners automate newly learned L2 segmental articulations via the development of new pronunciation *habits*.

UNDERSTANDING SEGMENTAL PRONUNCIATION AUTOMATICITY

The process of learning an L1 involves not only "tuning" the mind to hear some phonetic distinctions as phonemic and others as not, but also tuning the vocal organs to default to L1 configurations. As a result, students beginning to hear the difference between members of an L2 minimal pair and to be able to articulate new L2 phoneme(s) nevertheless often fail at pronouncing "learned" L2 phonemes accurately except in the rare instances when they consciously attend to pronunciation. As explained earlier, this is because pronunciation more than other language skills is necessarily nearly always produced automatically (i.e., subconsciously) since the mind does not have the processing capacity necessary to support the multiple higher-level cognitive processes required for fluent speech production as well as to simultaneously also control pronunciation consciously.

L1-trained pronunciations are so deeply entrenched that successfully replacing them in the stream of speech with more accurate L2 pronunciations can be the hardest step of all in L2 pronunciation acquisition. The almost magnetic nature of L1-based articulatory habits makes them very hard for L2 learners to break. Additionally, unless L2 learners are true beginners, their battle to entrench new L2 segmental pronunciation habits will include to a large extent working to *un*learn previously acquired (i.e., now fossilized) mispronunciations of particular L2 words containing their problem phonemes. Students may additionally have habits of mispronouncing certain words due to interference from L1 sound-spelling correspondences (Young-Scholten & Archibald, 2000).

It is therefore vitally important teachers realize segmental training on *how* a sound is physically articulated is not enough to accomplish successful L2 segmental acquisition. Just as activities promoting the development of accurate L2 sound *perception* are vitally important to accurate L2 sound *production*, activities promoting accurate L2 sound production as a *habit* are mandatory. L2 learners' pronunciation will only be fluently accurate in the stream of speech if they can 1) *hear* L2 phonemes accurately, 2) *articulate* them accurately and 3) accomplish both tasks subconsciously. We therefore recommend the training sequence described below to maximize L2 learners' success in developing and automatizing new, more accurate segmental pronunciations.

A RECOMMENDED TRAINING SEQUENCE

In light of the three key realities impacting L2 segmental pronunciation acquisition described above, to be maximally effective, segmental pronunciation teaching must involve the following:

1) Developing the accuracy of students' L2 sound perception categorization (including the ability to discriminate L2 contrasts not existing in their L1)
2) Developing students' awareness of *how* L2 sounds are articulated and building their ability to physically articulate these sounds
3) Helping students undo old pronunciation habits for problem sounds and automatize new ones

Recommended resources and assignments for accomplishing each step of this training sequence are described below.

Developing accurate L2 sound perception

To build students' ability to hear the difference between various L2 English segmentals, we recommend two tools designed to build learners' proficiency in distinguishing minimal pairs their L1 may have trained them to hear as allophones. Linguatorium Auris (Qian, Chukharev-Hudalainen, & Levis, 2018) diagnoses each student's L2 English perception difficulties, adapting its recommended 10 minutes of daily High-Variability Phonetic Training (HVPT) homework to continually focus on the student's current problem segmentals. Linguatorium Auris' online version as well as app interface are transparent and easy-to-use for both students and teachers and its pricing is adaptable to different lengths of school term (with all proceeds dedicated to continued maintenance and research development by its parent nonprofit, the Andrey A. Hudyakov Center for Linguistic Research). Another valuable HVPT tool available for free online and as an iOS (but not Android) app is English Accent Coach (Thomson, 2012), though it requires students to learn the International Phonetic Alphabet (IPA) as well as to manually adapt its games to reflect their step-by-step segmental training progress (Sheppard, 2016).

Ideally, only when these tools indicate learners have receptively mastered a particular phoneme distinction should they begin being taught how to configure their vocal organs in order to *produce* that distinction. However, since the mind's process of breaking L1 allophones into separate L2 phonemes is not instantaneous but instead takes place apparently incrementally over time, we have found that if learners reach an apparent impasse at some intermediate point in the process of mastering a particular phoneme distinction receptively, taking several days off from perception training (presumably providing their minds time to consolidate what they have already learned) or even moving ahead to metaphonological training as well as physical articulation practice can help clarify to students what in their perception training they should be listening *for* and often provide the breakthrough needed. In other words, according to our experience, the old chicken-and-egg conundrum applied to pronunciation pedagogy – "Which comes first – perception or production?" – is misleading because complete mastery of new L2 phonemes in terms of either perception or production appears not to occur at a single point in time, but rather to develop incrementally.

A useful consciousness-raising tool for helping students notice their nonstandard pronunciation of particular words is YouGlish. Students asked to identify how their pronunciation differs from five or more YouGlish pronunciations of the U.S. state of "Illinois," for example, are often able to identify the point of difference accurately (e.g., YouGlish speakers do not usually pronounce "Illinois" with a final [z]). Not only that, but because of this exposure to many YouGlish speakers whose shared pronunciation for a given word differs from their own, students are often highly motivated to adjust individual word pronunciations toward a more standard form.

Developing accurate L2 sound production

To develop students' conscious understanding – or metaphonological awareness – of *how* a target segmental is physically articulated, a useful website for *assessing* whether students' inaccurate pronunciation results from lack of explicit knowledge about *how* to articulate a particular segmental is Daniel Currie Hall's Interactive Sagittal Section. A useful website for *teaching* students how the various English (as well as Spanish and German) segmentals are physically articulated is the University of Iowa's Sounds of Speech.

#10	#9	#8	#7	#6	#5	#4	#3	#2	#1
/b, p/ **/f, p/** **/m, n/** **/l, n/** **/l, r/**	**/f, h/** **/d, t/** **/g, k/**	/v, w/ /s, z/	/b, v/ /f, v/ /ð, z/ /s, ʃ/	/ð, v/ /s, ʒ/	*/θ, ð/* */s, θ/* */d, ð/* */z, dʒ/* */n, ŋ/*	*/t, θ/*	*/tʃ, dʒ/*	*/ʃ, tʃ/* */ʃ, ʒ/* */j, ʒ/*	*/f, θ/* */j, dʒ/*

Highest functional load = **bold**
Lower functional load = regular
Lowest functional load = *italic*

Figure 2. Relative importance of English consonant pairs based on functional load (Brown, 1988; see also Catford, 1987; Jenkins, 2000, 2002; Munro & Derwing, 2006).

Because of the difficulty of breaking one's L1-trained pronunciation defaults, it is important students focus on only a few problem segmentals at any one time. The high-functional-load segmentals most responsible for distinguishing utterances should, of course, be prioritized over those that are low functional load, as exemplified in Figure 2 (Brown, 1988, 1991; Catford, 1987; Jenkins, 2000, 2002; King, 1967; Munro & Derwing, 2006).

Developing accurate L2 sound production automaticity

To develop students' segmental pronunciation automaticity, students can be provided adequate practice with new L2 segmentals via a series of activities that are at first relatively controlled but increasingly become less controlled and more authentic. Students who work just 10-15 minutes per day on building new L2 segmental pronunciation habits are likely soon to find their old L1-based habits no longer sound quite right to them and their new, more standard L2 pronunciations increasingly do.

A good initial tool for facilitating students' repeated practice of newly learned L2 segmentals and/or minimal pair contrasts toward the goal of entrenching more standard vocal configuration habits is Nilsen & Nilsen's (2010) *Pronunciation Contrasts in English*. A useful consciousness-raising tool for facilitating learners' noticing of *which* English words contain a newly learned or habitually mispronounced phoneme is *RelateWorldwide's North American English Pronunciation Highlighter*. Students simply:

1) Copy-and-paste into the Pronunciation Highlighter from a PowerPoint or Prezi presentation (or any other source text)
2) Indicate which problem segmentals they want highlighted (again, no more than 2-3 at a time is best since learners are limited in how many new segmental articulations they can focus on at any one time)
3) Optionally specify the HTML color code they prefer for highlighting each segmental
4) Click "Submit."

The *Pronunciation Highlighter* then outputs the text, providing lists of potential problem words broken down according to problem phoneme as well as displaying the entire text with all words containing one or more of the student's problem phonemes appropriately color-coded. Only when students can read through each phoneme's wordlist fluently as well as accurately should they practice the entire highlighted text, paying attention to pronouncing the problem phonemes in each highlighted word accurately.

For students in the habit of mispronouncing technical terms in their fields, either segmentally or in terms of lexical stress, a useful technique for identifying their nonstandard technical term pronunciations is having them systematically read aloud through the glossary of an undergraduate textbook introducing their field. For each listed term, students should:

1) pronounce the term,
2) embed the term in a sentence and
3) pronounce the term again.

Their teacher, meanwhile, should take notes on any pronunciation issues to be able to provide students individual instruction and tailored homework later. Students can similarly read aloud through conference or seminar presentation slides/posters to obtain their teacher's help in identifying general and technical vocabulary mispronunciations they characteristically use when discussing their specific research niche. This can be followed up by students repeatedly talking through one slide at a time of a PowerPoint/Prezi presentation, paying careful attention to their pronunciation of words (and particularly technical terms) they have historically mispronounced, as identified by RelateWorldwide's *North American English Pronunciation Highlighter*.

Once students are capable of articulating the standard pronunciation of a word but struggle to do so reliably, a useful method for "resetting" learners' pronunciation of a particular problem word is having them repeat the word in a thought group context for 10-15 new YouGlish examples every day for a week, paying careful attention to pronouncing each accurately (e.g., https://youglish.com/search/live vs. https://youglish.com/search/leave). By the end of this period, learners are likely immediately to recognize if they revert to their former mispronunciation and be able to correct it. Depending on how frequently students use a given problem word, in a very short time this can lead to their relatively completely recalibrating historically problematic word and ultimately phoneme pronunciations. YouGlish is thus a powerful tool for building L2 learners' fluency in *consistently* applying their new L2 word and segmental pronunciations to the stream of speech in real-life communication contexts.

CONCLUSION

To be maximally effective, L2 pronunciation instructors must work *with* rather than *against* the cognitive processes every human uses to understand receptively and speak productively any languages they know. A key reason L2 learners struggle to pronounce new segmentals is because their L1 has trained them to hear L2 phonemes as allophonic (Best & Tyler, 2007; Broersma & Cutler, 2008; Flege, 1995; Jenkins, 2000; Qian, Chukharev-Hudalainen, & Levis, 2018; Richards, 2012). Only if learners can reliably accurately *hear* L2 phoneme differences as phonemic are they likely to be able reliably accurately to self-assess their pronunciation of these phonemes. Only if learners can reliably accurately self-assess their L2 pronunciation without consciously attending to the task are they likely to make progress toward *reliably* accurately configuring their vocal organs to distinguish challenging L2 phoneme distinctions in their real-life spoken communication. Certainly, it is important that L2 pronunciation instructors put focus on learners' mouths in their segmental pronunciation teaching. However, it is equally important that instructors take into account learners' minds along with the habitual nature of segmental pronunciation. Only then can optimal L2 segmental pronunciation training take place.

ABOUT THE AUTHORS

Monica Richards is a Postdoctoral Fellow in the Center for Communication Excellence (CCE) at Iowa State University. In this role, she has set up the CCE's English Speaking Consultation program as well as developed extensive materials on best practice in the teaching of L2 English pragmatics, listening, speaking, pronunciation and lexicogrammar for training new English Speaking Consultants. Monica is passionate about research-grounded L2 pedagogy, materials development and teacher training. Monica also cares deeply about building L2 English researchers' self-editing skills so their submissions to international English-medium journals are judged solely in terms of research quality rather than (lack of) adherence to English research writing norms.

Elena Cotos is an Assistant Professor in the Applied Linguistics Program at Iowa State University. She is also the Director of the Center for Communication Excellence of the Graduate College at Iowa State University. She investigates computer-assisted language learning and assessment, English for specific and academic purposes, functional language in instructional discourse, corpus and genre-based pedagogy, genre writing in the disciplines and automated writing evaluation.

REFERENCES

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* (pp. 13-34). Amsterdam: John Benjamins.

Broersma, M., & Cutler, A. (2008). Phantom word activation in L2. *System, 36*(1), 22-34. https://doi.org/10.1016/j.system.2007.11.003

Brown, A. (1988). Functional load and the teaching of pronunciation. *TESOL Quarterly, 22,* 593-606.

Brown, A. (1991). Functional load and the teaching of pronunciation. In A. Brown (Ed.), *Teaching English pronunciation: A book of readings* (pp. 211-224). London, UK: Routledge.

Catford, J. C. (1987). Phonetics and the teaching of pronunciation: A systemic description of English phonology. In J. Morley (Ed.), *Current perspectives on pronunciation: Practice anchored in theory* (pp. 86-99). Washington, DC: TESOL.

Detey, S., & Nespoulous, J. L. (2008). Can orthography influence second language syllabic segmentation? Japanese epenthetic vowels and French consonantal clusters. *Lingua, 118*(1), 66-81.

Erdener, V. D., & Burnham, D. K. (2005). The role of audiovisual speech and orthographic information in nonnative speech production. *Language Learning, 55*(2), 191-228.

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). Timonium, MD: York Press.

Ingels, S. (2010). The effects of self-monitoring strategy use on the pronunciation of learners of English. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 1st Pronunciation in Second Language Learning and Teaching Conference*, Sept. 2009. (pp. 67-89). Ames, IA: Iowa State University.

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

King, R.D. (1967). Functional load and sound change. *Language, 43,* 831–852.

Munro, M. J., & Derwing, T. M. (2006). The functional load principle in ESL pronunciation instruction: An exploratory study. *System, 34,* 520-531.

Moran, S., McCloy, D., & Wright, R. (Eds.). (2014). PHOIBLE Online. Retrieved November 28, 2018, from http://phoible.org

Nilsen, D. L. F., & Nilsen, A. P. (2010). *Pronunciation contrasts in English* (2nd ed.). Long Grove, IL: Waveland Press, Inc.

Qian, M., Chukharev-Hudilainen, E., & Levis, J. (2018). A system for adaptive high-variability segmental perceptual training: Implementation, effectiveness, transfer. *Language Learning & Technology, 22*(1), 69-96. https://doi.org/10125/44582

Richards, M. (2012). Helping Chinese learners distinguish English /l/ and /n/. In. J. Levis & K. LeVelle (Eds.), *Proceedings of the 3rd Pronunciation in Second Language Learning and Teaching Conference,* Sept. 2011. (pp. 161-167). Ames, IA: Iowa State University.

Scovel, T. (1969). Foreign accents, language acquisition, and cerebral dominance. *Language Learning*, *19*(3–4), 245–253. https://doi.org/10.1111/j.1467-1770.1969.tb00466.x

Sheppard, B. (2016). English Accent Coach. *TESL-EJ*, *20*(1), Available online: http://www.tesl-ej.org/wordpress/issues/volume20/ej77/ej77m2/

Thomson, R. I. (2012). Improving L2 listeners' perception of English vowels: A computer-mediated approach. *Language Learning, 62*(4), 1231-1258.

Vaux, B. (2000, January). *Flapping in English*. Paper presented at the meeting of the Linguistic Society of America, Chicago, IL.

Young-Scholten, M., & Archibald, J. (2000). Second language syllable structure. In J. Archibald (Ed.), *Second language acquisition and linguistic theory* (pp. 64-97). Oxford, England: Blackwell.

Wallace, L., & Fernandez, S. (2019). Teaching Tip: Personalizing peak vowel training in stressed syllabus: A sneak peek at blue canoe for perception and production. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Ames, IA, September 2018 (pp. 423-433). Ames, IA: Iowa State University.

TEACHING TIP

PERSONALIZING PEAK VOWEL TRAINING IN STRESSED SYLLABLES: A SNEAK PEEK AT *BLUE CANOE* FOR PERCEPTION AND PRODUCTION

Lara Wallace, Ohio University
Sofía Fernandez, Ohio University

Placing stress on a different syllable or using a different vowel can momentarily confuse listeners or lead to listeners' miscomprehension. Because there are nearly three times the number of vowel sounds as letters for writing vowels in English, learning these sounds can be challenging. To distinguish between them, the Color Vowel Chart gives us a name for each vowel sound so that learners can understand, for example, that the peak vowel is pronounced "purple shirt" CIRcular, not "red pepper" SECular. Placing lexical stress accurately and reducing such vowel errors that may be of high functional load can help ELLs speak more intelligibly. In this teaching tip, we will have a look at a method for identifying lexical stress and peak vowel quality, then training learners' perceptions of vowel quality and practicing their production in context. This multimodal method is based on a communicative framework and utilizes the Color Vowel Chart and some of the features of *Blue Canoe,* an app that has been designed around the Color Vowel Chart.

INTRODUCTION

For English Language Learners (ELLs) to communicate effectively, their speech must be intelligible (Derwing & Munro, 1997). Having an accent is not necessarily linked with noncomprehension (Munro & Derwing, 1999); rather, when speakers misplace the stress on a syllable, listeners can be confused, and this act leads to possible miscomprehension (Bond, 1999; Cutler & Clifton, 1984). This kind of mistake can damage the understanding between individuals, and even though pronunciation errors can be mitigated by context, listeners must be able to understand what the speaker tries to convey (Field, 2005).

One key to understanding speakers is their clear pronunciation of stress as a suprasodic feature. Stressing part of a word or phrase "causes it to stand out from other unstressed elements" (Derwing & Munro, 2015, p. 59). As Gilbert (2008) explains, in each word, there is one syllable that is stressed more than the others, the peak syllable; she adds that the vowel quality of the peak syllable must be recognizable at the word level and the phrase (or thought group) level. Research on listeners' perceptions of intelligibility of lexical stress supports this notion (Cutler & Clifton, 1984; Field, 2005), encouraging the teaching of lexical stress patterns due to their communicative value (Field, 2005). In the following example from Gilbert's (1993) *Clear Speech,* we can see the potential for misunderstanding the word "committee" if someone stresses the first syllable instead of the second syllable, which also changes the vowel quality: "What did you think of the com**mit**tee? What did you think of the **co**medy?" (p. 69). In order to pronounce these words (and phrases) intelligibly, one's stress placement should be clear and the stressed vowel should be a bit longer than the other vowels. Thus, there is value in teaching English Language Learners (ELLs) lexical stress patterns (Field, 2005), and vowel quality of peak vowels where needed for successful communication.

In terms of vowel quality, ELLs might have a doubly difficult time identifying the quality of the peak vowel, however, since some segmentals will be absent from their L1's vowel inventory and since the many vowel sounds are represented differently in English orthography. Referring back to the previous example, the letter "o" is pronounced differently in the words "committee," "comedy," and "comb." Similarly, the letter "i" is pronounced in a number of ways, and will sound different whether full or reduced, as in the "i" in "live" in the instances of "to live" and "a live show." Thus, there is a need to help ELLs to train their perception, and research suggests that teaching segmentals can help learners address these potential barriers to intelligibility (Derwing, Munro, & Wiebe, 1998). One way to train ELLs' perceptions of vowel quality is by using the Color Vowel Approach (Taylor, Thompson, & Barr, 2016).

In this approach, colors are superimposed onto the IPA vowel quadrilateral so that each color represents a vowel sound. In this Color Vowel Chart, the phonetic symbol /ɪ/ is "silver pin" and /aɪ/ is "white tie," for example (Taylor, Thompson, & Barr, 2016, p. 5). In this way, ELLs have more than just an IPA symbol to help them understand the vowel quality—they also can begin associating vowel quality with color when they hear the sound repeated in the color vowel sound name and symbol. So, returning to the previous examples of the letters "i" and "o" above, instead of learning that the vowels are /ɪ/ live and /aɪ/ live, or /ə/ committee, /a/ comedy, /o/ comb, ELLs can perhaps gain a more concrete understanding of the differences in vowel quality both visually and aurally using the Color Vowel Chart. In these words, the "colors" of the "i" vowels are "silver pin" to live and "white tie" live, whereas those "o" vowels are "cup of mustard" committee, "olive sock" comedy, "rose boat" comb.

Taking it a step further, ELLs can organize words by vowel quality/color; for words with more than one syllable, the word can be classified by the quality/color of its peak vowel. To do so, ELLs can work with a dictionary to identify the syllables that receive primary stress in words, and use the Color Vowel Chart to identify the quality—or color—of the peak vowel. In the case of the word "committee," the stress is on the second syllable, so the color of the word is "silver pin" committee, whereas in "comedy," the stress is on the first syllable with the vowel /a/; therefore, its color is "olive sock /a/" comedy. Other "silver pin /ɪ/" words include "minute," "forgive," and "different," and other "olive sock /a/" words include "college" and "compartment." Thus, ELLs can learn peak vowel quality and perhaps even some word stress patterns when they learn "to speak in color" (Taylor, Thompson, & Barr, 2016, p. 7).

Learning the color vowel approach can be particularly helpful for ELLs who use different stress patterns in their speech than with what their audience may be familiar. After all, there is variation in the vowel inventory and lexical stress patterns from dialect to dialect. The latter is what can cause potential confusion. Take for example NAE's (North American English) "garAGE" versus British English's "GARage" or NAE's "ROtate" versus British English's "roTATE" (Celce-Murcia et al., 2010, p. 455). When a different stress pattern is used, it can be difficult for listeners to understand what was said, especially when the peak vowel's quality is different, as in "committee" and "comedy" (Cutler & Clifton, 1984; Field, 2005). Returning to the Color Vowel Approach to highlight the difference in vowel quality, the colors of each word respectively are "silver pin" committee and "olive sock" comedy. One such misunderstanding took place in an oral communication class at Ohio University where ELLs interact with domestic students. The domestic student was quite confused about what the ELL was explaining since she thought that he had said "CIRcumstance" when he had actually said "circumference." It seemed that the confusion arose because instead of stressing the second syllable (cirCUMference), he stressed the first: "CIRcumference." Once we had worked out what happened, the student found that he had to stress (what is known on the Color Vowel Chart as) "cup of mustard /ə/ cirCUMference" instead of placing the stress on the first syllable. With this explanation, he was easily able to adjust the stress pattern. Working on the pronunciation of certain words or sets of words that follow a pattern should help ELLs to speak more intelligibly since "the more words the listener is able to accurately identify, the more intelligible the speaker is" (Isaacs, 2008, p. 557). Thus, following lexical stress patterns assists in listener comprehension.

Part of intelligibility, as Kang and Moran's (2014) research suggests, requires these speakers to have few vowel errors that are of high functional load. Functional load refers to "the importance of linguistic phenomena in distinguishing meanings in a language" (Derwing & Munro, 2015, p. 74). For example, Kang and Moran (2014) illustrate low functional load as when speakers pronounce "adventure" as "advunture" (p. 180). In this case, they explain, the listener would likely understand the speaker's intention. By contrast, with a word that carries a higher functional load, as in the case of /ɪ/ vs. /i/ "live" instead of "leeve" (leave) (Kang & Moran, 2014, p. 179), the authors imply that listeners might have a more difficult time interpreting the message. In the case of word stress more specifically, they explain, intelligibility may be compromised if ELLs stress the second syllable instead of the first as in "visITing" instead of "VISiting" (Kang & Moran, 2014, p. 180), even though the vowel quality of the stressed syllable is the same (both syllables can be pronounced as "silver pin" /ɪ/). This difficulty in understanding a word when the stress shifts to the right is also evidenced in Cutler and Clifton's (1984) study on word recognition where they experimented with listeners' perceptions of words where lexical stress was shifted.

In order to help ELLs pronounce key words effectively in spontaneous and planned speech where word stress is a barrier to intelligibility, we propose the utilization of a multimodal method based on a communicative framework (Celce-Murcia et al., 2010) that uses the Color Vowel Chart and its Blue Canoe app to help ELLs improve their pronunciation. *Blue Canoe* is an app that utilizes the color vowel approach to practice listening discrimination and production of peak vowel quality. The app contains an introduction to the color vowel system along with instructional videos, several games, and its own dictionary that ELLs can use to look up the vowel quality of the syllable that receives primary stress—or as we will call it for the remainder of this paper, the color of a given word's peak vowel [i] For the purposes of this teaching tip, we will use the app's dictionary and Merriam Webster for help with analysis.

PROCESS

Step 1: Identify target words

The first step to personalized peak vowel training is to conduct a needs analysis in order to generate a word list for students to practice. These words should be a part of the students' regular lexicon and should consist of words and short phrases that may be difficult for the native listener to understand.

Generating the list is a two-fold process. The first step is to perform an initial assessment, and the second is an ongoing assessment. In this first assessment, ELLs answer a variety of Test of Spoken English-like speaking prompts[ii]. Since the end goal is for ELLs to communicate effectively in spontaneous speech, the needs assessment should therefore include free speech rather than merely reading a text or a word list (Levis & Barriuso, 2012). The rater (most likely the instructor) will listen to the speaking samples and make note of words and phrases that are difficult to understand due to a different word stress pattern (for example, if the speaker stressed the last syllable of "determine" instead of the second syllable). A careful listening of these words should give the rater an idea of whether the difficulty in understanding can be attributed to the consonant (probably not word stress), the number of syllables, which syllable receives the stress (if any), and the vowel quality of the stressed syllable. The level of detail provided in the feedback from the analysis would be up to the instructor, but at the very least, the analysis should generate a list of words for students to learn. This early assessment forms the basis of each learner's list but is just an initial snapshot that necessitates ongoing monitoring for other words and phrases.

The instructor and learners should add to each individual list throughout the semester. This means that the instructor takes notes any time learners give a presentation and speak up in class. The learners, in turn, should monitor their everyday speech interactions. For example, when listeners stop and ask for clarification or seem confused, learners make note of their particularly confusing pronunciation. To generate and maintain this list, instructors and ELLs can collaborate on a Google doc (or something similar that both parties can add to). In keeping with the focus of this activity, the words and phrases that we will highlight are those with lexical stress and vowel quality that need to be clear in order to avoid misunderstandings, such as cirCUMference, which can be confused with CIRcumstance if the stress placement is placed on the first syllable, or CIRcular, which can be confused with SECular if the vowel quality is not sufficiently distinct. Figure 1 below is part of a sample list.

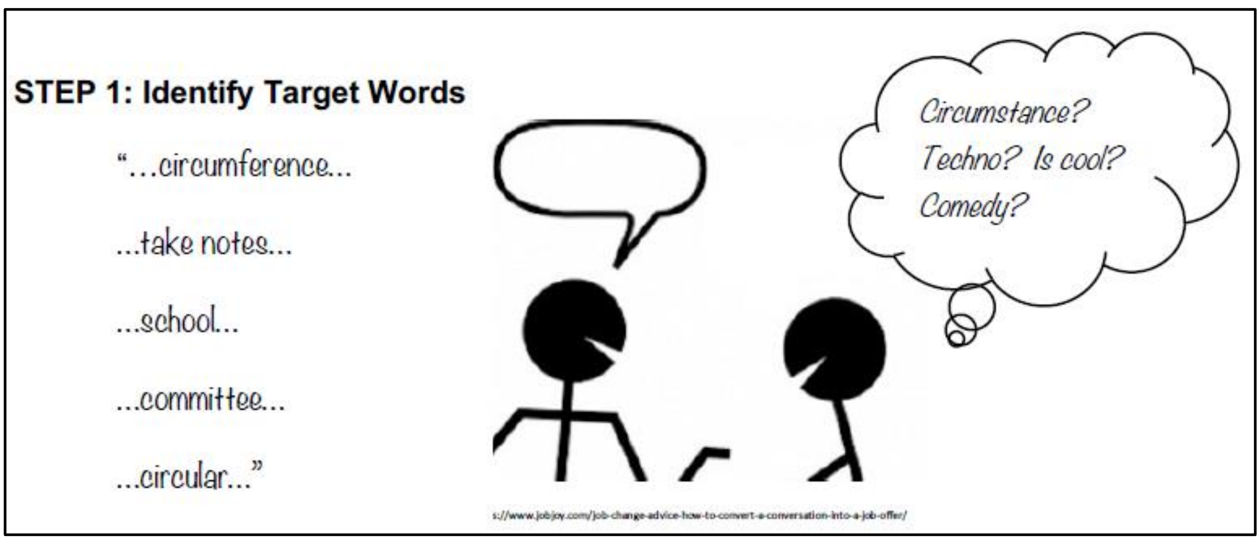
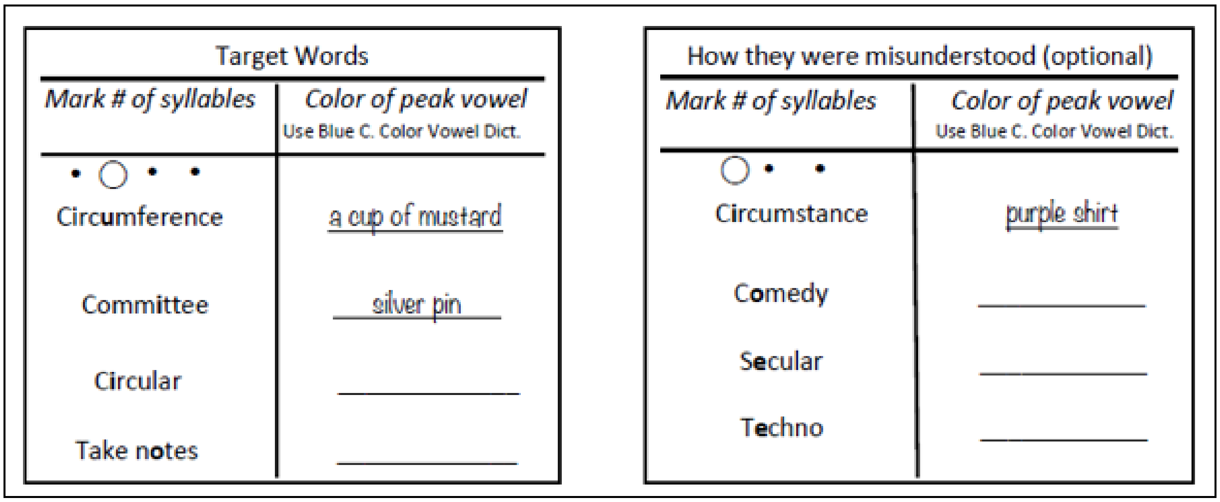


Figure 1. Identify target words to create the word list.

Step 2: Learning word stress patterns and identifying peak vowel quality: Analysis, exploration, and exercises in perception

Once the learners' target words have been identified, they should then listen to the word to determine the number of syllables and the vowel quality/color of the peak vowel in the target dialect. Focusing on such characteristics can lead to improved intelligibility (Derwing, Munro, & Wiebe, 1998). Two options for doing so include: 1) as a class activity where the instructor or classmates model the word; or 2) as self-study by looking it up in the Blue Canoe app's Color Vowel Dictionary; or if they do not have the app, by listening to how the word is pronounced and seeing how it is transcribed in Merriam Webster's dictionary online.

To help students understand the number of syllables, instructors can use Judy Gilbert's kinesthetic technique of tapping out the syllables to check the number of syllables (Gilbert, 1993, p. 1). This activity lends itself well to full class participation since students can often be reluctant to tap out syllables when working on their own. To help students understand counting syllables visually, they can mark the stressed syllable with a large circle, and unstressed syllables with smaller dots (see Figure 2 for an illustration).



Target Words	
Mark # of syllables	*Color of peak vowel* Use Blue C. Color Vowel Dict.
• ○ • • Circumference	a cup of mustard
Committee	silver pin
Circular	____
Take notes	____

How they were misunderstood (optional)	
Mark # of syllables	*Color of peak vowel* Use Blue C. Color Vowel Dict.
○ • • Circumstance	purple shirt
Comedy	____
Secular	____
Techno	____

Figure 2. Analyze word stress patterns and identify the color of the peak vowel.

To identify the peak vowel, students must first become aware of the qualities of a stressed syllable. These stressed syllables are typically longer in length, often are higher in pitch, and are perceived as louder (Fry, 1958). When there are multisyllabic words that exhibit secondary stress, these syllables are often longer in length but not necessarily higher in pitch. To simplify this analysis for the students, focus on the syllable with primary stress and have students underline the peak vowel.

To figure out the vowel quality of the stressed syllable, students should match the sound to the color vowel. For instance, the peak vowel in "circumference" is "cup of mustard /ʌ/" cirCUMference, and the peak vowel in "CIRcumstance" is "purple shirt /ɚ/" CIRcumstance. If the student has the Blue Canoe app[iii], students can use the dictionary in the app (see Figure 3) to check their work. They can also listen to the recording of the word in this dictionary to determine the number of syllables.

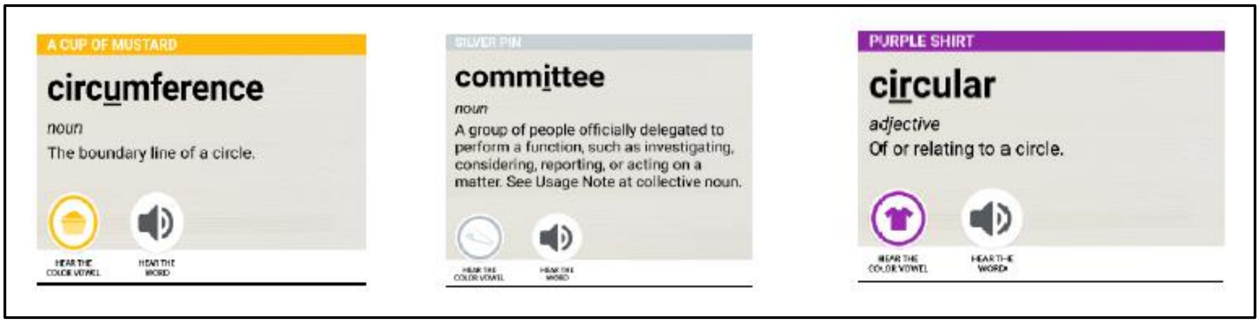


Figure 3. Screenshots from the color vowel dictionary in the Blue Canoe App.

Students can then organize their words and phrases by vowel quality of the peak vowels using the Color Vowel Organizer (see Figure 4)[iv]. This Organizer uses the corresponding vowel colors and icons, serving as a visual reminder to the students of what the vowel sounds like.

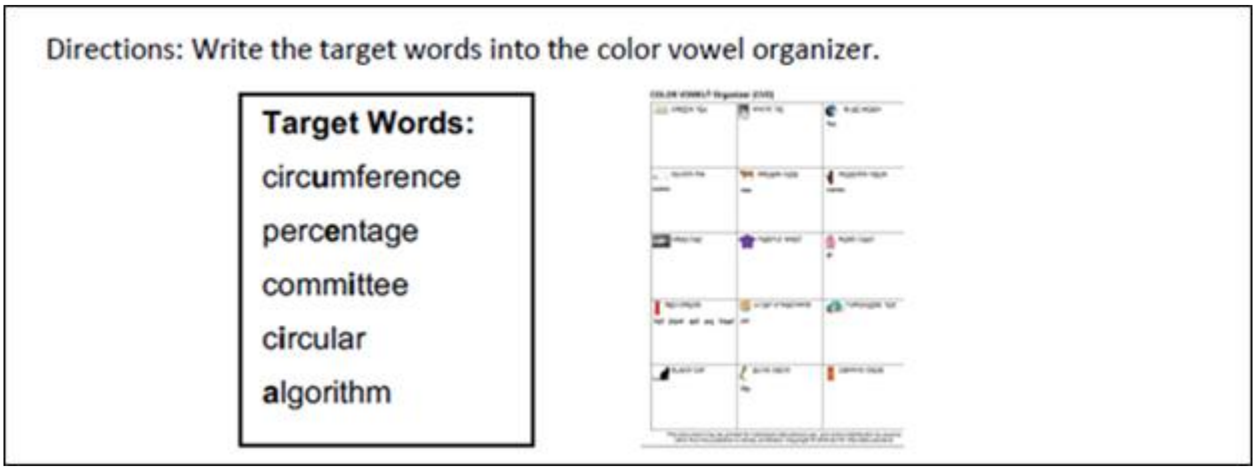


Figure 4. Target words & color vowel organizer.

For more perception practice, learners can plug each word into YouGlish to hear how the word is pronounced in the target dialect (as the site broadly classifies videos into US, UK, or Australian English) by slowing down the video. Students can also scroll down to look at the "nearby words" and make note of any similar sounding words to their target words; minimal pairs are particularly useful for helping students understand the need for pronouncing each word in the pair in a distinct way (for example, "committee" and "comedy").[v]

Step 3: Controlled production practice with the color vowel organizer

Once the Organizer is filled out, students should check their classification by color/vowel quality (with others or alone) by saying the color, sound, then the word as illustrated in Figure 5. If the pronunciation of the peak vowel is different from the color and sound, the student should reconsider the pronunciation or the placement of the word on the Organizer. To facilitate and perhaps strengthen students' self-monitoring skills, the students should record themselves and listen back. They can compare their pronunciation to that of Merriam Webster's dictionary. The instructor or another perceptive individual can also provide feedback.

"Red pepper /E/perCENtage"
"Black cat /ae/ Algorithm"
"Olive sock /a/ biOlogy"
"Red pepper /E/ proFESsor"

Figure 5. Peak vowel classification and how to read the words in the context of their color.

Once the student confirms the correct classification of the words, in that all peak vowels should have the same vowel quality, it is time to practice. This can be done as a class or individually. The students will work vowel by vowel, again saying the color, sound, and word or phrase. This technique is known as "flooding" and is encouraged by Taylor, Thompson, and Barr (2016) for the purpose of helping the students to "notice the vowel quality" (p. 9). Through these controlled practice activities, students can gain a better understanding of vowel quality and number of syllables.

Step 4: Guided production practice

For this step, students create their own Color It Out game (which is both a physical card game and a game in the Blue Canoe app) to play with their classmates by writing the words from their personalized word list onto color vowel-coded index cards. They can do this at home or in class, or if necessary, the instructor can make this set (putting each student's name on the bottom of each of their cards). See Figure 6 for what each card might look like.

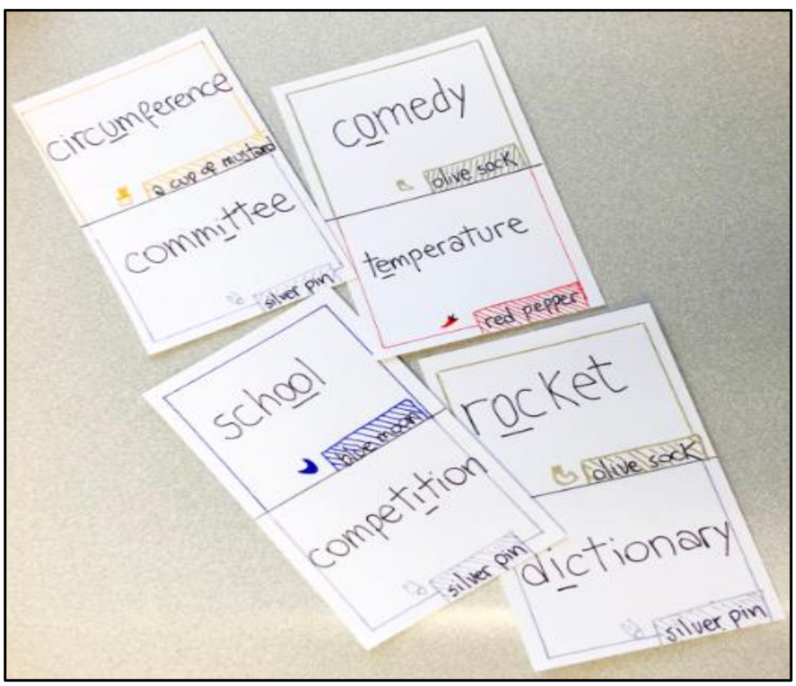


Figure 6. Personalized Color It Out cards.

Once the cards are created, the students can play the game in pairs or small groups after the instructor models the rules. If the students are using the Blue Canoe app, they may have already played the game and will know how to play. If not, here is how to play a short game. Pair up the students, dealing out four cards per player, then designate the rest of the cards as a draw pile. Turn a card from the draw pile face up. Of the two colors listed on the card, the first player looks at their hand and finds a card with one of the matching colors, places it next to the face-up card, and reads the color pair, listening for whether they are saying the vowel in the same way and stressing the word accurately. For example (see Fig. 7), they would say “green tea TEAcher, green tea Even.” (Please note that Figure 7 illustrates [Color it Out] as it looks in the app, but learners will use their own cards for this step.) The next player will look at the top card and find a matching color, and this will continue. If a player does not have a matching color in their hand, they can draw a card. The game is over when a person runs out of cards. Instructors might stress to the students that the goal is not to win, per se; rather, the goal is to practice.

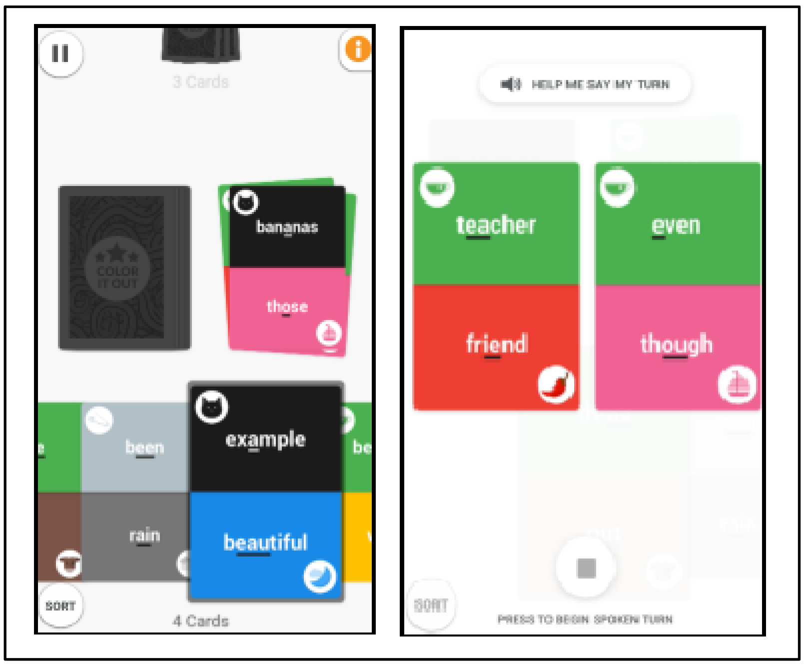


Figure 7. Screenshots from Blue Canoe’s Color It Out game set up.

There are a few ways for students to add to their perceptual practice. One is by recording their turns. Students can record their turns to listen back and/or send it to the instructor to check. This is slightly different than the app where students cannot listen to their own speech afterwards[vi]; instead, the voice recognition technology in the app decides if the vowel quality of the stressed syllable was sufficiently accurate. The other way is listening to a model speak their turn first. With the personalized handmade game that we describe here, students can ask someone else--like their partner or an instructor--to say the turn first. This is similar to a feature in the app where the player can click "Help Me Say My Turn" to hear what it should sound like (as shown in Figure 7 on the top-right).

Step 5: Communicative production practice

This last step prompts learners to practice their words in focused free speech. In other words, they will intentionally use the words in context. The first way they can do so is to define and explain each term, making sure to emphasize it when using it as in the following example:

> "CirCUMference is the distance around a circle. Let's take for example a cake. A cake is usually round, so if we know the cirCUMference of the cake, we will have a good idea of how many servings there are. To find the cirCUMference of the cake, simply…"

Another way to practice these words in context is to create a short dialogue that uses the word as illustrated below.

> CirCUMference
> A: How do you find the cirCUMference of this table?
> B: To get the cirCUMmference of the table, multiply Pi by the diameter of the circle.
> A: Under what CIRcumstances do we need to calculate the cirCUMference of something...?

Regardless of the ways that the words are put into practice, learners should record their production so that they can listen back, notice their peak vowel production of the word in question, and analyze their speech, then try again if the peak vowel or stress placement is unintelligible. Once they understand how to produce the words intelligibly, it is important to continue to practice mindfully and to self-monitor. With enough mindful application and repetition, it should be possible for the ELL to move from conscious to unconscious competence.

LEARNERS' FEEDBACK, CONCLUSION

After piloting this teaching tip and the use of *Blue Canoe* as part of improving peak vowel production and stress placement with a small number of graduate student ELLs, we discovered through one-on-one interviews a number of benefits and a few areas of caution. ELLs consistently reported the usefulness of being able to think of the words on their lists as a color. One person illustrated this when she noted that she thinks of words as colors; she explained that "disease is green tea--I can see the underline [of the stressed syllable], the color, and the image of green tea." From the instructor's end as well, this color vowel approach helps to give clear feedback on ELLs' production. For example, instructors are able to point out when something is a "silver pin" word rather than a "green tea" word.

Another useful item of note to the ELLs was the color vowel dictionary. Although the voice is computer generated, they reported that it was useful to look up a word ("algorithm," for example) and see what color the peak vowel was (e.g., "black cat /æ/" algorithm). As a potential solution in response to the artificial voice and to hear it in connected speech, ELLs were also encouraged to use YouGlish (US search) to listen to these words in context.

It is our hope that through this multimodal process, instructors will have a clear and engaging way to help ELLs improve their peak vowel pronunciation.

ABOUT THE AUTHORS

Dr. Lara Wallace is a Senior Lecturer and the Coordinator of ELIP's Pronunciation & Presentation Lab in ELIP Global & Academic Communication Program, a part of the Department of Linguistics at Ohio University. She has taught international teaching assistants for more than ten years and oral communication for 20 years. Her research interests include pronunciation pedagogy to improve International Teaching Assistants' intelligibility, CALL, pronunciation pedagogy, student motivation for learning, and empowering education.

Contact information:
383 Gordy Hall
1 Ohio University Drive
Athens, OH 45701
Phone: (740) 593-1991
Email at wallacl1@ohio.edu

Sofia Fernandez is a 2019 graduate of the MA program in Linguistics and was the Assistant Coordinator of ELIP's Pronunciation & Presentation Lab at Ohio University. She holds an MA in Spanish Linguistics and Literature, a TEFL certificate, and a CALL certificate. Her research interests include second language pronunciation and bilingualism.
Email: fernandez.sofia10@gmail.com

REFERENCES

Bond, Z. (1999). *Slips of the ear: Errors in the perception of casual conversation*. San Diego, CA: Academic Press.

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

Cutler, A. & Clifton, C. E. (1984). The use of prosodic information in word recognition. In H. Bo,uma & D. G. Bouwhuis (Eds.), *Attention and performance X* (pp. 183-196). Hillsdale, NJ: Lawrence Erlbaum.

Derwing, T., & Munro, M. (1997). Accent, intelligibility, and comprehensibility: Evidence from L1s. *Studies in Second Language Acquisition, 19*(1), 1-16.

Derwing, M. & Munro, M. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. Amsterdam; Philadelphia: John Benjamins Publishing Company.

Derwing, M., Munro, M., & Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning, 48*, 393-410.

Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly, 39*(3), 399-423.

Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech 1*, 126-152.

Gilbert, J. (1993). *Clear speech: Pronunciation and listening comprehension in North American English* (2nd ed.). Cambridge, UK: Cambridge University Press.

Gilbert, J. (2008). *Teaching pronunciation: Using the prosody pyramid*. Cambridge, UK: Cambridge University Press.

Isaacs, T. (2008). Towards defining a valid assessment criterion of pronunciation proficiency in non-native English-speaking graduate students. *The Canadian Modern Language Review, 64*(4), 555-580.

Kang, O., & Moran, M. (2014). Functional loads of pronunciation features in nonnative speakers' oral assessment. *TESOL Quarterly*, *48*(1), 176-186. http://doi.wiley.com/10.1002/tesq.152

Levis, J., & Barriuso, T. A. (2012). *Nonnative speakers' production errors in spoken and read English*. In J. Levis & K. LeVelle (Eds.), Proceedings of the 3rd Annual Pronunciation in Second Language Learning and Teaching Conference (pp. 187-194). Ames, IA: Iowa State University.

Munro, M., & Derwing, T. (1999). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning, 49 (Supp. 1),* 285-310.

Taylor, K., Thompson, S., & Barr, R. (2016). *The Color Vowel approach: Resources for connecting pronunciation to vocabulary, reading, and spelling*. Santa Fe, NM: ELTS.

---

[i] Interested educators can contact Blue Canoe (partners@bluecanoelearning.com) about accessing the app since at the time of publication, the teacher is integral to the successful use of the app; thus, it is not for individual download.

[ii] For sample questions, see pp. 7-11 of the pdf here https://www.ets.org/s/toeic/pdf/speaking-writing-sample-tests.pdf

[iii] Again, interested educators can contact Blue Canoe (partners@bluecanoelearning.com) about accessing the app.

[iv] The Color Vowel Chart and Color Vowel Organizer are available for purchase on the Color Vowel Launchpad, and available at https://elts.solutions/product/cv-launch-pad/ .

[v] For more on how to use YouGlish, see Wallace, L. (2018, March). Using YouGlish with ITAs to increase input and speaking opportunities. *ITAIS Newsletter*. Retrieved from: http://newsmanager.commpartners.com/tesolitais/issues/2018-03-07/4.html

[vi] It is possible for players to listen and compare their speech to the model's speech in *Let's Talk*, another game in the Blue Canoe app that at the time of writing is in Beta mode.

Becker, K., & Edalatishams, I. (2019). ELSA Speak – Accent Reduction [Review]. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Ames, IA, September 2018 (pp. 434-438). Ames, IA: Iowa State University.

TECHNOLOGY REVIEW

ELSA Speak - Accent Reduction

Kimberly Becker, Iowa State University
Idée Edalatishams, Iowa State University

INTRODUCTION

ELSA Speak - Accent Reduction is an application (app) for reducing a non-native English accent. *ELSA* (an acronym for "English Language Speech Assistant") has a free version and a paid-for version for Android and Apple products. *ELSA*'s paid version can be purchased for $3.99 monthly, $8.99 for three months, or $29.99 for one year. As described on their website, this app employs "proprietary artificial intelligence" (AI) including automatic speech recognition (ASR) to provide feedback on users' pronunciation accuracy. This review will provide an overview of the basic features before moving to a critical evaluation of the free version of the app.

Description

ELSA's basic interface is simple and provides navigation between topics and skills, levels, reports, and other features. Users can set a display language (English, Vietnamese, or Japanese – good indicators of the target audience for this app). Depending on the proficiency level of the user, the skills include exercises such as beginning and ending sounds, minimal pairs, schwa, *th*-sounds, and consonant clusters. Applied linguists typically refer to these as segmental phonemes—individual sounds within a word. The topics include food, entertainment, technology, culture, daily conversations, and relationships, among many others. Figure 1 demonstrates the default skills-based page, which appears after the user chooses their proficiency level.

Figure 1. The skills-based start-up page.

After choosing a skill (e.g., "/S/ /SH/" as shown in Figure 2), users can select a topic, under which they can choose from several lessons. Each lesson is labeled for proficiency level and includes various speaking and listening opportunities. Figure 3 illustrates an exercise at the *regular* level with the target sound underlined.

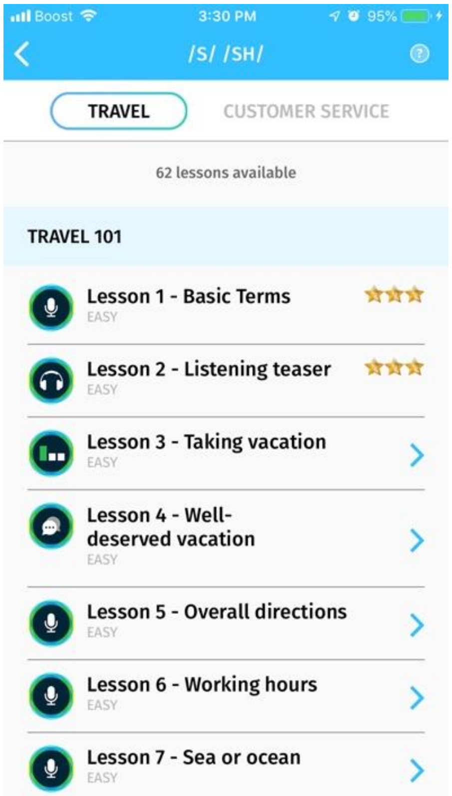


Figure 2. Lessons for the /S/ /SH/ "skill".

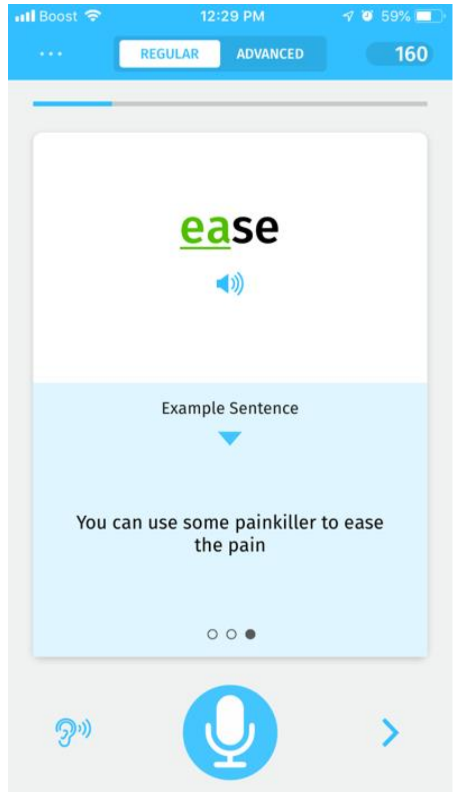


Figure 3. An example activity at the *regular* level.

Users listen to the pronunciation of the target sound and then record themselves. *ELSA* determines the accuracy of the recording and provides feedback for correct and incorrect pronunciations. Correct pronunciation elicits a bell sound accompanying a score up to 100; incorrect pronunciation elicits a buzzing sound along with the correct pronunciation of the sound or word and an opportunity to review the recorded incorrect sound for comparison purposes. An example of feedback given for a user's incorrect pronunciation can be seen in Figure 4.

Besides organizing by skill, *ELSA* also provides "Topics" in the bottom menu, as shown in Figure 5.

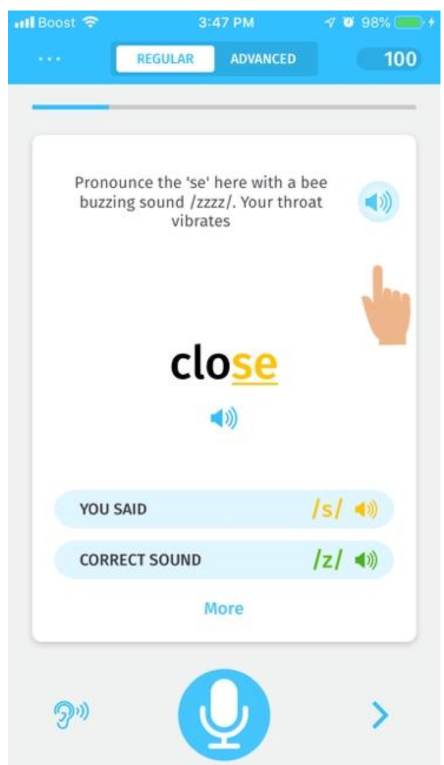


Figure 4. Incorrect pronunciation feedback.

Figure 5. The topic-based page.

The app also includes a dictionary that provides the pronunciation of any word plus the opportunity to connect to Youglish.com, a website offering YouTube clips with audiovisual emphasis on particular sounds, words, or phrases. *ELSA* also offers a multi-faceted feature called "Progress" containing a "Word Bank," "*ELSA* Pronunciation Score" (EPS), and "Assessment." "Word Bank" offers a progress summary in three sections: "Word Sound," "Word Stress," and "Conversation." Word sound specifies feedback about individual phonemes and word stress assesses syllable stress.

The second progress feature, EPS, is a collection of scores from *ELSA* activities. The third progress feature, the "Assessment," allows users to record 13 sentences and then gives feedback on segmental issues such as consonant clusters, aspiration sounds, and schwa. *ELSA* provides a percentage correct score, and users can also get a detailed report. Figure 6 illustrates an example of a detailed report.

The green letters indicate that the user correctly pronounced the sound, the yellow letters indicate that the sound was mostly correct, and the red letters means that the sound was incorrectly pronounced. Despite the fact that the test is marketed as a way to measure "speaking proficiency level" and prompts users to produce sentences, feedback is still focused on individual segmental sounds.

Figure 6. An example of detailed report.

Besides these features, *ELSA* offers a user "Profile," where users can customize settings about feedback, sharing, and notifications. Additionally, in the profile, users can identify goals, set a daily reminder, and view the app's terms and policies.

EVALUATION

As claimed on its website, *ELSA* has been featured on several digital media websites and magazines and has 4.5/5 ratings from 23,000 users. However, this pronunciation app could improve in many ways. This section will provide a critical evaluation of this pronunciation app in relation to best practices for Mobile Assisted Language Learning (MALL), using Reinders and Pegrum's (2016) framework.

One major shortcoming is that *ELSA* focuses only on segmental aspects of pronunciation. The spotlight on individual sounds is not an issue in and of itself; however, the lack of balance between segmental and suprasegmental practice opportunities is problematic. Pronunciation research literature shows that focusing pronunciation instruction on suprasegmental features improves comprehensibility and fluency more than segmental focus (e.g., Derwing & Rossiter, 2003; Kang, Rubin, & Pickering, 2010). In fact, Kang et al. (2010) note that "listeners can tolerate a great deal of inaccuracy in pronouncing consonants and vowels," (p. 555) as long as suprasegmental features are somewhat accurate. Other research also demonstrates the value of suprasegmental features (e.g., intonation and speaking rate) over segmental features (Anderson-Hsieh, Johnson, & Koehler, 1992; Derwing & Munro, 1997). While two of the categories in Reinders and Pegrum's (2016) framework for MALL app evaluation relate app design to pedagogical approaches, *ELSA* has almost no focus on suprasegmental features that are likely to be the most beneficial in accent reduction.

Another category in Reinders and Pegrum's (2016) framework is the potential for educational affordances. *ELSA*'s developers have misrepresented the abilities of their proprietary AI system, which often mistakenly identifies incorrect sounds as correct. The documentation does not provide any information about its accuracy or piloting. Moreover, *ELSA* seems to have been developed with a focus on quantity over quality. This app has hundreds, if not thousands, of individual exercises about segmental phonemes, but with an inaccurate system for catching mispronunciations combined with inattention to suprasegmental features, it is not likely to be effective for English language learning experts or discerning learners. Even with the paid version of the app, users only get access to more lessons; it does not offer a substantially better version except in providing extra exercises about the same skills and topics.

Correspondence of app design to principles of second language acquisition (SLA) is another category in Reinders and Pegrum's (2016) framework. In a meta-analysis of studies in second language pronunciation instruction, Lee, Jang, and Plonsky (2014) found that an important aspect among many of the studies was the inclusion of both segmentals and suprasegmentals. Lee et al. (2014) further suggest three important aspects of SLA as related to pronunciation: (1) using segmental and suprasegmental approaches, (2) aligning lessons with needs analyses, and (3) considering demographic information such as learner backgrounds and/or first languages. *ELSA* implements neither a suprasegmental approach nor any needs analysis. Because users can input information about their language background in their user profiles, it can be argued that the third of these aspects has been considered in *ELSA*'s design, although it is not clear how the designation of a first language affects the interaction between the user and the app.

The final category of the MALL evaluation framework is related to affect, which Reinders and Pegrum (2016) define as engagement and attention to affective filter. As accurate evaluation of the *ELSA* app in this category would require assessments from non-native speaking users, it will not be discussed here but would be an appropriate consideration for further development of the app.

Several of these issues can be at least partially explained by the fact that the *ELSA* development team does not include any applied linguists or other English language learning experts. Although the 11 employees featured on the app's website have backgrounds in speech processing, software development, and engineering, a linguistics or an SLA perspective is necessary to make this app more useful for the audience.

Finally, a seemingly surface-level (but important) criticism about the *ELSA* Corporation website and app are the multiple typos. These include but are not limited to missing words, pluralized noncount nouns (e.g., "feedbacks"), and confusing if not conflicting references to numbers and other statistics without reference to any sources. One typo that stood out as particularly egregious is the misspelling of the word *diphthong [sic]* on the assessment page.

CONCLUSION

ELSA is an app that has made great strides in the world of AI for practicing the pronunciation of individual sounds. However, in order for it to break new ground in the world of accent reduction, it needs to be expanded to include suprasegmental aspects of pronunciation. The lack of attention to what many applied linguists consider the most important part of accent reduction combined with smaller issues such as typos cause the app to lose face with the population of experts (language teachers) who are best positioned to both evaluate and market it.

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.

Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition*, *19*(1), 1-16.

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.

Kang, O., Rubin, D. O. N., & Pickering, L. (2010). Suprasegmental measures of accentedness and judgments of language learner proficiency in oral English. *The Modern Language Journal*, *94*(4), 554-566.

Lee, J., Jang, J., & Plonsky, L. (2014). The effectiveness of second language pronunciation instruction: A meta-analysis. *Applied Linguistics*, *36*(3), 345-366.

Reinders, H., & Pegrum, M. (2016). Supporting language learning on the move: An evaluative framework for mobile language learning resources. In B. Tomlinson (Ed.), *Research and materials development for language learning* (pp. 219-232). New York, NY: Routledge.

Elliott, T., & Baghestani, S. (2019). Sounds: The Pronunciation App [Review]. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Ames, IA, September 2018 (pp. 439-443). Ames, IA: Iowa State University.

TECHNOLOGY REVIEW

Sounds: The Pronunciation App

Tom Elliott, Iowa State University
Shireen Baghestani, Iowa State University

INTRODUCTION

Computer assisted pronunciation teaching (CAPT) has the potential to provide students with high quality pronunciation instruction inside or outside the language classroom (Levis, 2007). The benefits of CAPT include the ability to provide "individualized instruction, frequent practice through listening discrimination and focused repetition exercises, and automatic visual support that demonstrates to learners how closely their own pronunciation approximates model utterances" (Levis, 2007, p. 184).

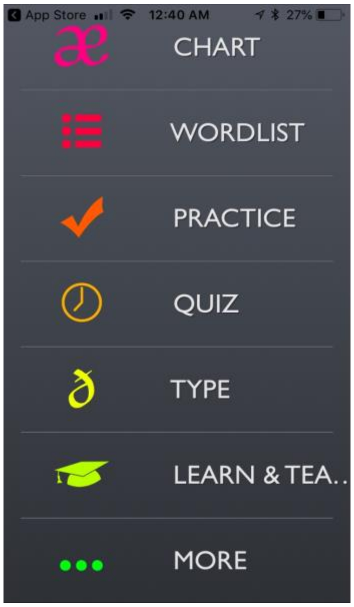
In addition, research has shown that L2 pronunciation may improve with the ability to match L2 phonemes to their corresponding IPA symbols (Lambacher et al., 2005), and CAPT technology is well-suited to provide this kind of practice. Macmillan Education's mobile application *Sounds: The Pronunciation App* is built around this principle, with activities that strengthen students' ability to match English phonemes with their corresponding IPA symbols.

DESCRIPTION

The application was first released in 2011 and was last updated in 2015. It is available for both iOS and Android OS systems (iOS Version 3.42 was used for this review). A free version with limited functionality is available, and the full version can be purchased for $5.99.

At the heart of the application is the Sound Foundations chart created by Adrian Underhill. This chart organizes the phonemes of English according to their articulatory properties (see Figure 1b and accompanying description). Users encounter this chart in various forms: they can interact with it as a standalone tool, and it serves as the keyboard for typing in IPA symbols.

The front page of the app lists six options: 'Chart,' 'Wordlist,' 'Practice,' 'Quiz,' 'Type,' 'Learn & Teach,' and 'More' (see Figure 1a). The 'Chart' option takes the user to an interactive version of the Sound Foundations chart (shown in Figure 1b). Here the user can tap a phoneme to hear how it is pronounced, or tap and hold to hear its pronunciation and an example word. There is also a 'Rotate for full screen view' function, but it did not work for the system tested (iPhone 6 iOS 11.4).



(a)

(b)

Figure 1(a). Front page of the application. *Figure 1(b).* Sound foundations chart.

Note: monophthongs, diphthongs, and consonants are grouped in separate quadrants of the chart (top left, top right, and bottom, respectively). Likewise, the position of each phoneme within each quadrant corresponds to the position of the tongue (e.g., /i/ requires having the tongue at the top front section of the mouth).

The second option from the main menu page is ‘Wordlist.’ From here the user is prompted to select a wordlist (see Figure 2a). The paid version of the application comes with both a General British and General American wordlist, each consisting of 650 words, with additional wordlists available for purchase ($0.99 each). Unfortunately, the British wordlist, once opened, caused the application to freeze and therefore was not usable. Figure 2b displays the American wordlist. Here, users can select a word to hear it pronounced, view its definition, as well as record and listen to their own pronunciation.

The third and fourth options from the main menu, ‘Practice’ and ‘Quiz’ are very similar. Both contain ‘Read,’ ‘Write,’ and ‘Listen’ modes. In the ‘Read’ mode a word is displayed in IPA notation and the user must write it in conventional English spelling before hearing it pronounced. ‘Write’ mode shows normal spelling and asks the user to write the word using IPA symbols. Finally, the ‘Listen’ mode plays a voice recording first, which users can replay as often as they like, and asks for an IPA transcription. Quiz mode includes the same exercises as the Practice mode, but gives a limited amount of time and a maximum number of allowed mistakes, as well as saving high scores.

(a)

(b)

Figure 2(a). Front page of 'Wordlists' mode. *Figure 2(b).* American word list.

The fifth option from the main menu is 'Type', which gives the user a phonetic keyboard and allows the text to be copied into other applications, such as email or text messages. The final option in the main menu is 'Learn and Teach', which provides extra resources for teachers and learners such as articles that discuss learning strategies and lesson plans that explain how to incorporate the app into classroom teaching.

EVALUATION

In this evaluation, we adopt Yoshida's (2018) four criteria for evaluating CAPT tools: 1) appropriateness to learning objectives, 2) quality and accuracy, 3) practicality of use, and 4) cost. In terms of the first criterion, the primary benefit of the application is in helping learners improve their ability to match sounds with their IPA symbols. This type of training can help students improve their pronunciation by teaching them to recognize when the same phoneme is used in different words, giving them a better understanding of the English sound system as a whole. Moreover, the application provides students with models of correct pronunciation, records their progress (e.g., in quizzes), and provides ample opportunities for independent practice (Yoshida, 2018). However, a major shortcoming of the application is that while it provides feedback on students' written input, it does not provide any feedback on their spoken input. Thus, while users of the application receive ample feedback for improving their production and recognition of IPA symbols, they must rely on their own ear to determine whether or not their spoken production matches the model pronunciation.

Another limitation for meeting users' learning objectives is the lack of focused or targeted practice. In the introductory video included in the app, phoneme chart creator Adrian Underhill asserts that, "There can't be a 'sound syllabus'. We have to have all of the sounds immediately." However, not all pronunciation errors are equally serious (Levis, 2007), so this approach of presenting learners with all phonemes at once may not be the most effective or efficient approach.

In terms of quality and accuracy, the application is generally effective. However, a few discrepancies are worth noting. For example, when the 'American English' option is selected, one still finds non-rhotic pronunciations in some of the audio, as well as British English vowels, such as /aɪ/, in the transcriptions. Also, there is inconsistency between audio recordings their IPA transcriptions. For instance, an American pronunciation may be provided in the audio form, but the written transcription reflects the British pronunciation.

In terms of practicality, the app is relatively easy to use, and because so many students are familiar with smartphones, it seems like a useful medium for increasing their outside practice time. However, one source of difficulty may be the fact that the IPA keyboard is laid out in the same way as the Sound Foundations chart. Although this chart is logically organized, users will need time to become familiar with the location of each phoneme, and this may be a source of frustration early on. Another limitation is that the system is unforgiving of spelling mistakes. Given that IPA spellings are not standardized (one can prove this by simply typing a word such as "ability" into various IPA transcription engines and observing the discrepancies), a user could produce a perfectly defensible phonetic spelling that nevertheless is counted wrong in the application.

Finally, the cost of the application ($5.99) is quite high given the limited functionality it provides. There are many educational apps available, and many websites that can provide similar features for free. Considering those factors, it seems unlikely that a student would pay so much for a phone app unless required to do so. Also, the lack of recent upgrades may dissuade users from spending the money since the app may soon become obsolete with changes in phone technologies and operating systems.

CONCLUSION

Sounds: The Pronunciation App is generally easy to use and provides an effective way of learning to match IPA symbols with their corresponding to English sounds. This type of practice can help to improve pronunciation by helping learners to recognize when the same phoneme is used in different words and with different orthography. It can also help learners become familiar and more comfortable with English phonemes. However, for $5.99 the app offers limited functionality when compared with other software that has been developed since its initial release. The learning outcomes are limited due to its focus on phonemes within isolated words, and it does not cover important pronunciation elements such as phonemic blending or prosody. For these reasons, it may be most suitable for learners who want to improve their pronunciation at the word level, or to learn to use the phonetic alphabet.

REFERENCES

Lambacher, S. G., Martens, W. L., Kakehi, K., Marasinghe, C. A., & Molholt, G. (2005). The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Applied Psycholinguistics, 26*, 227–247.

Levis, J. (2007). Computer technology in teaching and researching pronunciation. *Annual Review of Applied Linguistics*, *27*, 184-202.

Macmillan Publishers. (2018). Sounds: The pronunciation app. *Macmillan Apps.* Retrieved from http://www.macmillaneducationapps.com/soundspron/

Yoshida, M. T. (2018). Choosing technology tools to meet pronunciation teaching and learning goals. *CATESOL Journal*, *30*(1), 195-212.

Fattany, A., & Elnegahy, S. (2019). Sounds of Speech [Review]. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Ames, IA, September 2018 (pp. 444-448). Ames, IA: Iowa State University.

TECHNOLOGY REVIEW

Sounds of Speech 3.0

Afnan Fattany, Iowa State University
Sondoss Elnegahy, Iowa State University

INTRODUCTION

Recently, teaching pronunciation has received much attention and second language (L2) teachers have been interested to include it in their curriculum (Derwing, 2018). Fortunately, there are a variety of computer-assisted pronunciation training (CAPT) programs that are offered online, as a mobile application, or both. The question then becomes which one to choose. This paper reviews *Sounds of Speech 3.0,* a free online website that was developed for teaching phonetics for three languages: American English**,** German**,** and Spanish. The website also has a mobile application that can be downloaded on iOS or Android devices for $3.99. However, the application is only for American English pronunciation training, but provides translation for Korean, Spanish, and Chinese (Simplified and Traditional). This review will describe and evaluate the free website version of Sounds of Speech 3.0, referred to henceforth as SOS.

DESCRIPTION

The goal of the SOS website is to provide a thorough introduction to the segmental features of three languages: American English, German, and Spanish. To introduce these segmentals, a variety of tools are provided including, animated articulatory diagrams for each sound, annotated descriptions of how each sound is produced, a facial view video, and an audio sample of each sound.

SOS focuses on providing segmentals input in which users are provided with a categorization of sounds into consonants and vowels. Consonants are divided according to the manner of articulation, the place of articulation, and the voicing of the consonant. Vowels are divided into two types in American English and German, monophthongs and diphthongs, whereas, in the Spanish module, vowels are divided into semi-consonants, semi-vowels, and diphthongs. The user can select the manner, place, or voice of consonants or one of the vowel types to read a brief description to which it refers. Figure 1 shows this interface. The screenshot on the left shows that when "manner" is selected in the English consonants module, a description of the manner of articulation appears. The image on the right shows the detailed information when "stop" is selected.

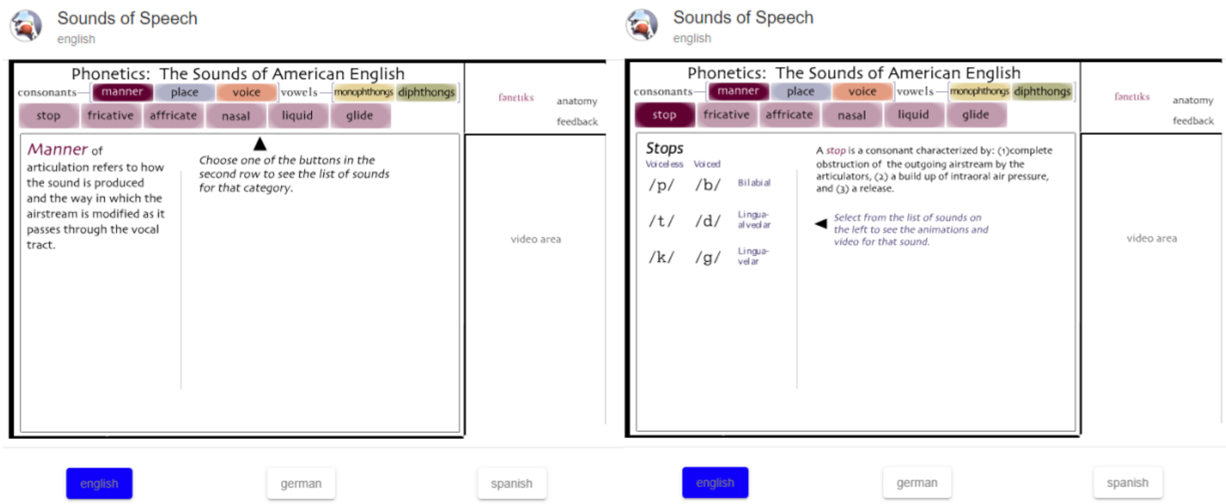


Figure 1. Selection of manner of consonants (left) and selection of the category stops (right) in the English Module.

Once a category is chosen, the phonetic transcription for each sound is provided, along with words that use the target segmental in word initial, medial, or final position (see Figure 2). In the Spanish model, these examples are provided in both the Spanish alphabet and phonetic transcriptions; however, for the American English and German modules, the words are only given in their phonetic transcription. Within the interface, users can play an animated articulatory diagram with the audio of the target sound by clicking the ‘animation with sound’ button at the bottom of the screen. Users can also follow a step-by-step annotation with a highlighted articulatory diagram of how to produce the sound by clicking the ‘step-by-step’ button, also at the bottom of the interface. In addition, users can watch a video with the front view of a native speaker’s lips pronouncing the sound located on the top right side of the screen. In the Spanish and German modules, users can play a video of the examples. In addition, videos include male and female native speakers of Spanish (with two different female speakers), but American English and German modules have only a female native speaker for audio and video.

Figure 2. Interface for the Spanish /b/ sound.

EVALUATION

To evaluate SOS, we are using guidelines developed by Neri, Cuchiarini, Strik, & Boves (2002) which center on how input, output, and feedback are used within the CAPT tool. Input refers to the quality and quantity of language received by a language learner, whereas output is the language produced by the learner. Finally, the different kinds of feedback given by teachers, peers, or native speakers can help learners improve their production.

According to Neri et al.'s (2002) guidelines for CAPT evaluation, SOS has successfully achieved its goal of offering users comprehensive input for the sounds of the three target languages. SOS is considered a good first step in learning the target language sounds in American English, German, and Spanish since the perception of language sounds is essential for accurate production (Neri et al., 2002). In addition, literature in L2 pronunciation teaching has supported the positive outcome of explicit phonetic instruction (e.g., Gordon, Darcy, & Ewert, 2013; Lord, 2005; Saito, 2011; Venkatagiri & Levis, 2007). This is clear in SOS by the detailed description of the segmental features and the explanation of how to produce the sounds via the animated articulatory diagram and the step-by-step annotation. In addition to providing input in a written form, SOS also presents the sounds in audio and audio-visual modes which helps learners to get an essential information about the different aspects of L2 pronunciation (Navarra & Soto-Faraco, 2007; Neri et al., 2002). The Spanish module presents the best example among the other two languages modules since it has a variety of input from three different native speakers of both genders.

Although users can hear each sound in isolation and in different position in words, the examples in SOS for all modules are only provided as individual words. Providing meaning context can play a vital role in improving pronunciation learning (Avery & Ehrlich, 1992; Neri et al., 2002) and this might be a future development for SOS to consider.

CONCLUSION

On the whole, *Sounds of Speech 3.0* is successful in achieving its goal in offering a comprehensive description of the segmental features of the target languages. SOS is also a good source of authentic input of segmental of the target language provided. In this sense, SOS can be a useful starting point for teachers in presenting segmental sounds of the target language and as perception training for L2 learners. Teachers can use the animated diagrams or videos along with the audio of native speakers for in-class instruction or to encourage learners to practice at home. Since the tool is free, it is accessible for students to use independently. Nevertheless, we recommend that teachers use SOS in classroom instructions while supplementing it with production activities and adequate feedback since these are integral parts of teaching and learning L2 pronunciation (Neri et al., 2002).

REFERENCES

Avery, P., & Ehrlich, S. (1992). *Teaching American English pronunciation*. New York, NY: Oxford University Press.

Derwing, M. T. (2018, September). Putting an accent on the positive: new directions for l2 pronunciation research and instruction, *Symposium conducted at the International Symposium on Applied Phonetics*, Aizuwakamatsu, Japan.

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*. Aug. 2012. Ames, IA: Iowa State University, 194-206.

Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly, 39*(3), 369-377.

Lord, G. (2005). (How) can we teach foreign language pronunciation? On the effects of a Spanish phonetics course. *Hispania*, *88*(3), 557-567.

Navarra, J., & Soto-Faraco, S. (2007). Hearing lips in a second language: Visual articulatory information enables the perception of second language sounds. *Psychological Research 71*, 4-12.

Neri, A., Cucchiarini, C., Strik, H., & Boves, L. (2002). The pedagogy-technology interface in computer assisted pronunciation training. *Computer Assisted Language Learning, 15*(5), 441-467.

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), 45-49.

Venkatagiri, H. S., & Levis, J. M. (2007). Phonological awareness and speech comprehensibility: An exploratory study. *Language Awareness*, *16*(4), 263-277.

Goodale, E., & Yang, J. (2019). English Pronunciation [Review]. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Ames, IA, September 2018 (pp. 449-453). Ames, IA: Iowa State University.

TECHNOLOGY REVIEW

English Pronunciation

Erik Goodale, Iowa State University
Junghun Yang, Iowa State University

INTRODUCTION

The widespread use of mobile technology has been changing lives since Apple's smartphone was introduced about ten years ago. Mobile devices provide unlimited access to online data from almost any location at any time. Thus, mobile devices have saturated many people's lives and have become a primary platform for information, entertainment, and social interaction. Following this change, the application of mobile technology to language education has been discussed by many researchers. A number of studies have reported that mobile application use is an effective way to learn a second language in that the multimodal features not only enable teachers to apply special instruction for resolving learners' individual differences in a classroom setting (Chen, Huang, & Wu, 2017), but the technology also creates an "immersive environment for the individual language learners" (Rosell, 2018, p. 871). Furthermore, the core users of mobile applications have high expectations for their mobile application as their only language learning source (Rosell, 2018). We would also suggest that it is important to scrutinize such applications from a linguistic perspective. In this review, we examine a mobile application called English Pronunciation which provides instructions for learning pronunciation in 30 languages and utilizes multimodal features and voice recording functions in order to provide language learners with an accessible environment that facilitates the development of their English pronunciation. We will provide a general overview of the featured segmentals, examine the application's multimodal characteristics, and evaluate the strengths and weaknesses of the application.

OVERVIEW OF MOBILE APPLICATION

English Pronunciation was developed by Awabe, a company located in Vietnam, which develops mobile applications for second language learners. Awabe claims that by using English Pronunciation, "You will learn some important topics about speaking English properly". These important topics consist of six categories of learning segments, namely short vowels, long vowels, double vowels (diphthongs), voiced consonants (such as /b/, /d/, and /g/), voiceless consonants (such as /p/, /t/, and /k/), and other consonants (such as /m/ and /n/). Figure 1 below shows the application's interface and demonstrates how the program addresses segmentals.

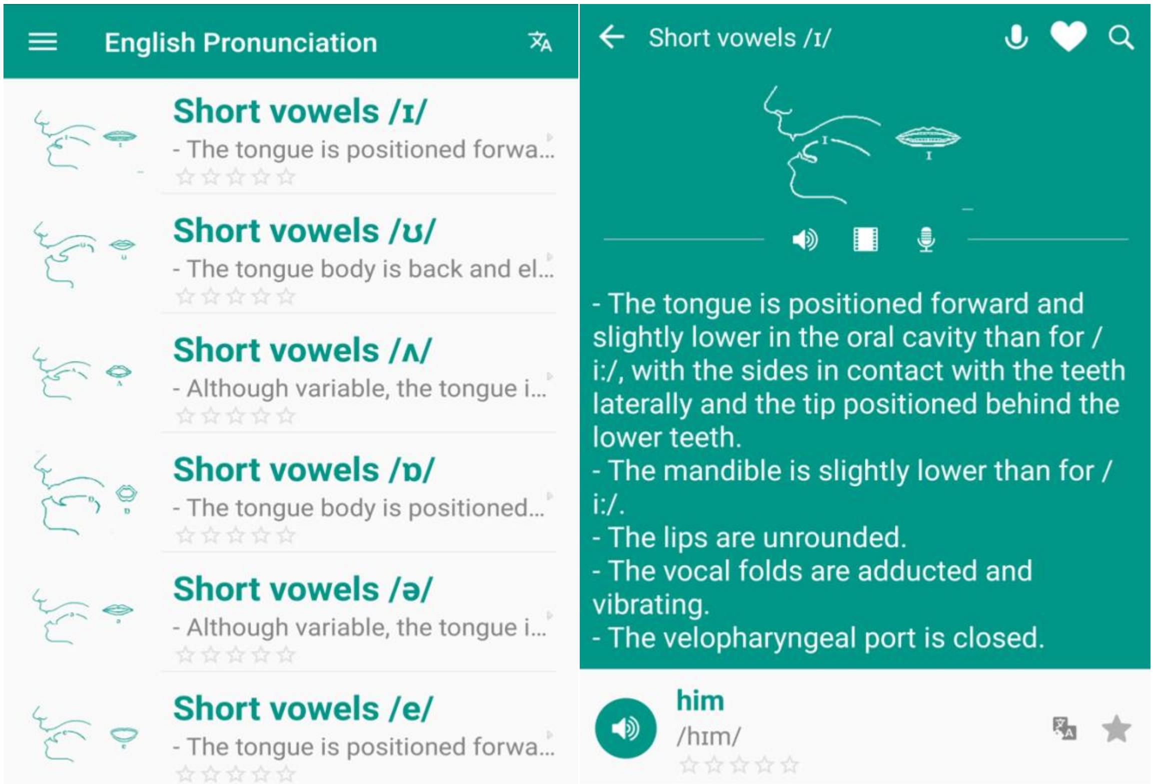


Figure 1. The interface of English Pronunciation.

Once users download the application, they will get notification from this application about training pronunciation such as “Today’s pronunciation” every morning. Users can modify the schedule of the notification in the application’s settings.

Multimodal Interaction

Each section provides multimodal resources for learning and practicing English pronunciation. Each vowel and consonant contain a picture showing lip rounding and a diagram of the human mouth to depict the position of the tongue inside of the mouth. Below the picture, links for audio and video content for the target pronunciation are provided with a description. This aligns with the category of description and analysis within the communicative framework for teaching pronunciation (Celce-Murcia, Brinton, Goodwin, & Griner, 2010).

Communicative Framework

After the learning process, the application provides a practice section that records the learners' pronunciation and plays back the recorded samples. Then it allows users to self-evaluate their pronunciation by selecting a rating out of a maximum of five stars. A screenshot of an example of this evaluation is presented in *Figure 2*. Standard American and British Received accents are given as target examples, and learners can record their pronunciation by clicking the recording button. These listen and repeat exercises can be used for both listening discrimination and controlled practice of the target segmental advocated in Celce-Murcia's Communicative Framework for Teaching Pronunciation (Celce-Murcia, Brinton, Goodwin, & Griner, 2010).

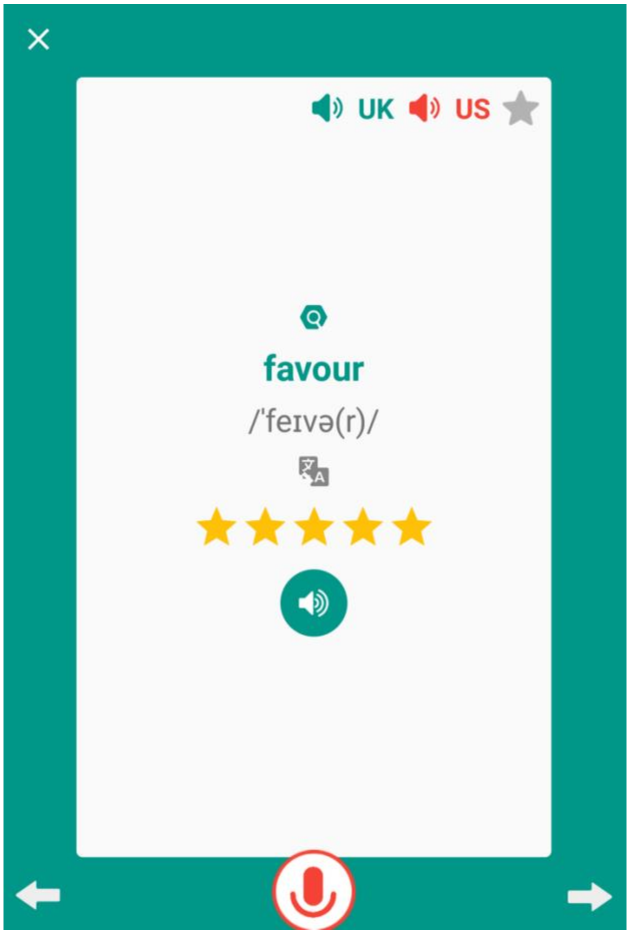


Figure 2. Self-evaluation of user pronunciation.

EVALUATION

The application has several merits for learners. This application is free so that learners can perform the pronunciation practice without cost. Instead, the application contains several five-second advertisements which appear when transitioning to different learning sections. Samples of the target phonemes commonly-used in English vocabulary are listed with audio content and phonemic description, and users can check them to make lists for later revision. In addition, this application's notification function can serve as a motivator that reminds users to practice their pronunciation. However, no guidance is given to users regarding prioritizing some sounds over others, and it is unclear what consideration was given to the concept of functional load. According to Derwing and Munro (2015), some sounds in English, such as /θ/, do not have as great of an effect on how comprehensible the speaker is, and thus errors involving these sounds have a lower level of gravity. Therefore, focusing on such sounds over those with a high functional load would be a less efficient use of time. The audio for each word is from a different speaker, which can be beneficial for learners to perceive different voices and is consistent with research on High Variability Phonetic

Training (HVPT) which posits that exposure to a variety of speakers promotes language acquisition (Bradlow, 2008). Furthermore, while the listen and repeat exercises are an important part of the communicative framework mentioned above, one downside is that the program uses them almost exclusively without incorporating other types of controlled practice activities such as minimal pairs, nor does it include guided practice and/or communicative practice activities also advocated by the framework. This might limit users' learning of pronunciation in a diverse context in real life (Celce-Murcia, Brinton, Goodwin, & Griner, 2010).

The descriptions of how English sounds are produced should be revised as they use terms which are difficult for language learners unfamiliar with technical terms related to the vocal tract to understand. This application also uses the terms "long" and "short" to define English vowels which can cause confusion. For example, some second language learners may understand the distinction between long and short vowel as lengthening of vowels, but other learners may understand the distinction as that a vowel letter is called "long" if it is pronounced the same as the letter's name (/aɪ/ in *twine*) and "short" if it is pronounced differently from the letter's name (/ɪ/ in *twin*), which is not part of linguistic terminology but commonly used in English classrooms especially when teaching native English-speaking children how to read vowels. The short vowels in this application consist of seven vowels, namely, /ɪ/ (as in *kit*), /ʊ/ (as in *foot*), /ʌ/ (as in *strut*), /ɒ/ (as in *lot*), /ə/ (as in the first syllable of *ago* and in the second of *sofa*), /e/ (as in *men*), and /æ/ (as in *trap*). Long vowels consist of /iː/ (as in *fleece*), /uː/ (as in *goose*), /ɔː/ (as in *thought*), /ɑː/ (as in *father*), and /ɜː/ (as in *nurse*). We found that this distinction between long and short vowels referred to as "received pronunciation" is defined in the Concise Oxford English Dictionary. However, this definition of long and short vowels is still problematic since the distinction seems to be based on whether the vowels are phonologically long or short, which is ambiguous in linguistics and does not address the vowel lengthening in different contexts such as the "short" vowel /ɪ/ in the word *'ridge' /rɪdʒ/* having a longer duration than "long" vowel /iː/ in '*reach*' */riːtʃ/.*

As a part of our evaluation, we also considered users' reviews. The English Pronunciation application seems to have already been used successfully according to the scores of around 6,900 users, which gave a score of 4.5 out of 5 in the spring of 2019. This is the average review score created by users and presented in Google Play Store that provides mobile applications, and digital content for Android mobile devices. However, we found that some of the users complained about the transcription using British pronunciation while the sample words were read by an American. Although this application provides multimodal features for pronunciation instruction, the discrepancy between phonemic description from British English and listening samples from American English can cause problems that confuse language learners getting incorrect information on the phonemic system of their target language.

CONCLUSION

English Pronunciation has interesting features as a pronunciation learning tool, such as a multimodal description of points of articulation and listen and repeat exercises that follow a part of the communicative framework in language learning. It is also free for users and has notification function encouraging users' practice. However, it has confusing terminology, relies on users evaluating themselves rather than providing them with external feedback, and does not include more communicative tasks.

One technological limitation of the mobile application is that it does not fully use programs such as voice recognition for analyzing learners' pronunciation instead of having users evaluate themselves. By using voice recognition, the application provides a helpful evaluation of learners' pronunciation to identify learners' weaknesses in their second language pronunciation that they might otherwise miss. Some mobile applications provide an evaluation of learners' pronunciation in a paid version although the accuracy and reliability of the evaluation are low according to reviewers. It is a very difficult process to program mobile applications to accurately recognize and analyze speech across different speakers and provide feedback as if it were a native speaker of the target language. The program also does not include features beyond vowels and consonants with aspects such as intonation, stress, and pitch being ignored. It would be beneficial for learners to have exercises utilizing these features as well and could be an area of future development. As this mobile application continues to develop and adopt voice recognition features, English Pronunciation could be established as a beneficial tool for second language learners.

REFERENCES

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 Publishing Company

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

Chen Hsieh, J. S., Huang, Y.-M., & Wu, W.-C. V. (2017). Technological acceptance of LINE in flipped EFL oral training. *Computers in Human Behavior, 70*, 178–190. https://doi.org/10.1016/j.chb.2016.12.066

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

English Pronunciation, Awabe (n.d.). Retrieved from https://play.google.com/store/apps/details?id=com.awabe.englishpronunciation

Rosell-Aguilar, F. (2018). Autonomous language learning through a mobile application: A user evaluation of the busuu app. *Computer Assisted Language Learning, 31*(8), 854–881. https://doi.org/10.1080/09588221.2018.1456465

Guskaroska, A., & Bordbarjavidi, F. (2019). LanguageTwo.com [Review]. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Ames, IA, September 2018 (pp. 454-460). Ames, IA: Iowa State University.

TECHNOLOGY REVIEW

LanguageTwo.com

Agata Guskaroska, Iowa State University
Fatemeh Bordbarjavidi, Iowa State University

INTRODUCTION

LanguageTwo.com, a website created by Thomas David Kehoe in January 2017, is intended to train second language learners in speech perception, segmentation, pronunciation, and diglossia, but the primary target is second language learners with auditory processing disorders, those who cannot "pick up a language by ear". The author argues that if the website can help learners with auditory processing disorders, then all second language learners will find the website helpful. Kehoe claims that the website primarily aims to improve the ability to hear a second language. The website is free and can be used on any device (e.g., computer, tablet, cell phone) with internet access and any type of software. According to the author, the website is still being developed, which needs to be taken into consideration when evaluating this website.

Currently the tool supports four languages (English, Spanish, Chinese, and Finnish), however the author does not provide reasons for selecting these four languages. The website is still under construction for Spanish, Chinese, and Finish with final improvements being made for the English part. For the purposes of this review, we focus on English (only American English is an option at this point).

To create an account, learners are encouraged to use their Google, Facebook, Twitter, or GitHub account; email is given as an option, but not recommended by the author.

DESCRIPTION

The main page of the website which provides learners with a menu consisting of different explanations, descriptions, and instruction for phonemes, consonants, vowels, and IPA (see Figure 1 below). The menu also shows IPA for the vowels (front, central, and back) and consonants (stops/plosives, affricates fricatives, sibilants approximants, and nasals). When learners click on one of the words, a written description appears as a pop up (e.g., English short I, *kit, lid, fill, bin*) including a description of the position of the sound in the mouth near-close, near-front, unrounded vowel are provided for learners. Learners can click on the headphones sign and hear the phoneme produced by a native American speaker.

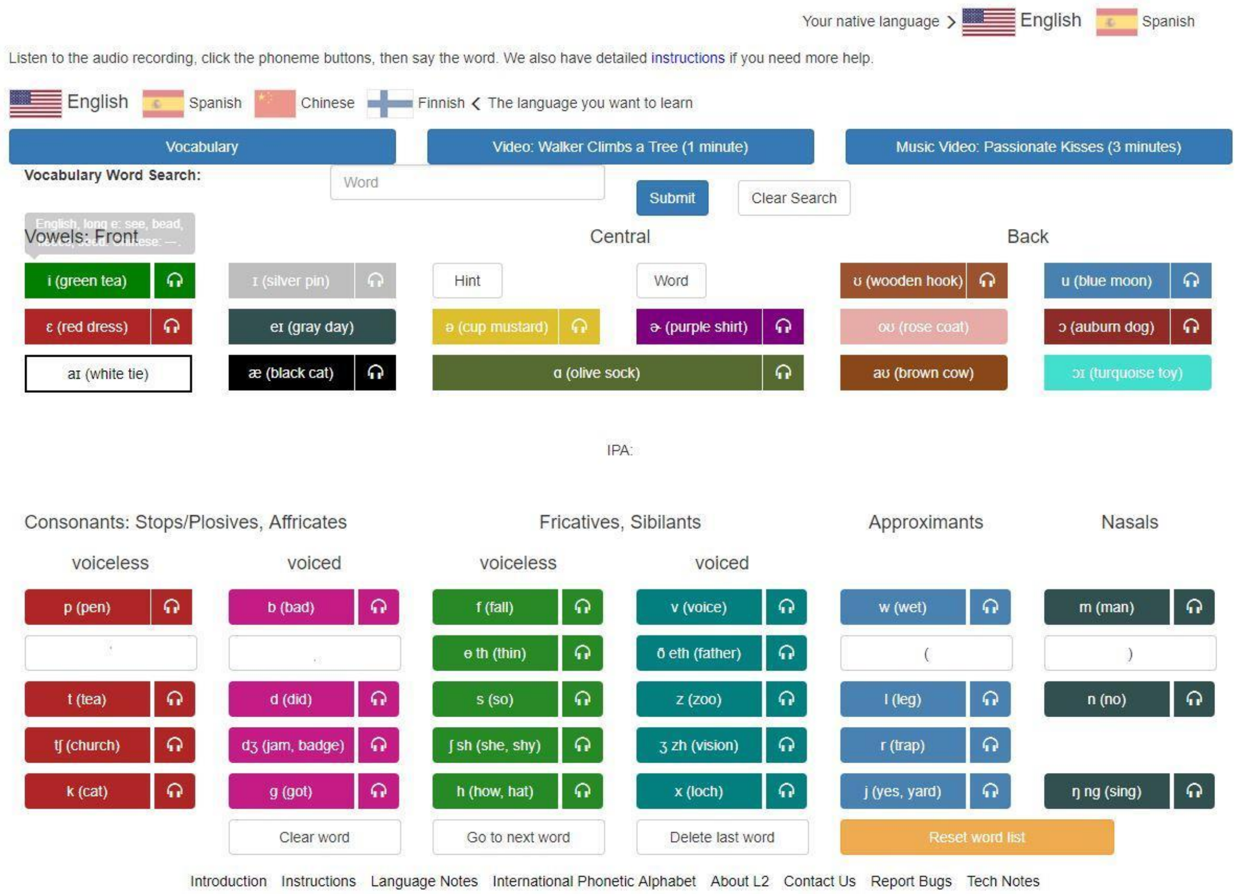


Figure 1. Main page of the website.

In addition to this main page, the website includes three sections: Vocabulary, Video, and Music Video (see the top of Figure 1). A brief description of the three main sections along with their screenshots are provided below.

Vocabulary

After clicking on "Vocabulary", users are provided with a variety of options. One such option is a search bar which allows users to search for any word being linked to the Oxford English Dictionary and pulls up the information on the LanguageTwo website (see Figure 2, section 1 below). The author chose to use this dictionary in order to include the word, IPA transcription of the pronunciation, audio recordings of exemplary speakers saying the word (they used synthesized speech) and definitions, translations, speech recognition (to test the learners' pronunciation of words) (T. Kehoe, personal communication, November 27, 2018).

Furthermore, learners can listen to the pronunciation of individual vowels and consonants, read a brief description and hear/see examples of each sound (See Figure 2, section 2). Then, learners can click on the provided sounds and try to click on the phoneme that they heard (Figure 2, section 3). Finally, there is also a hint button that lets learners find the correct vowel or consonant (Figure 2, section 4).

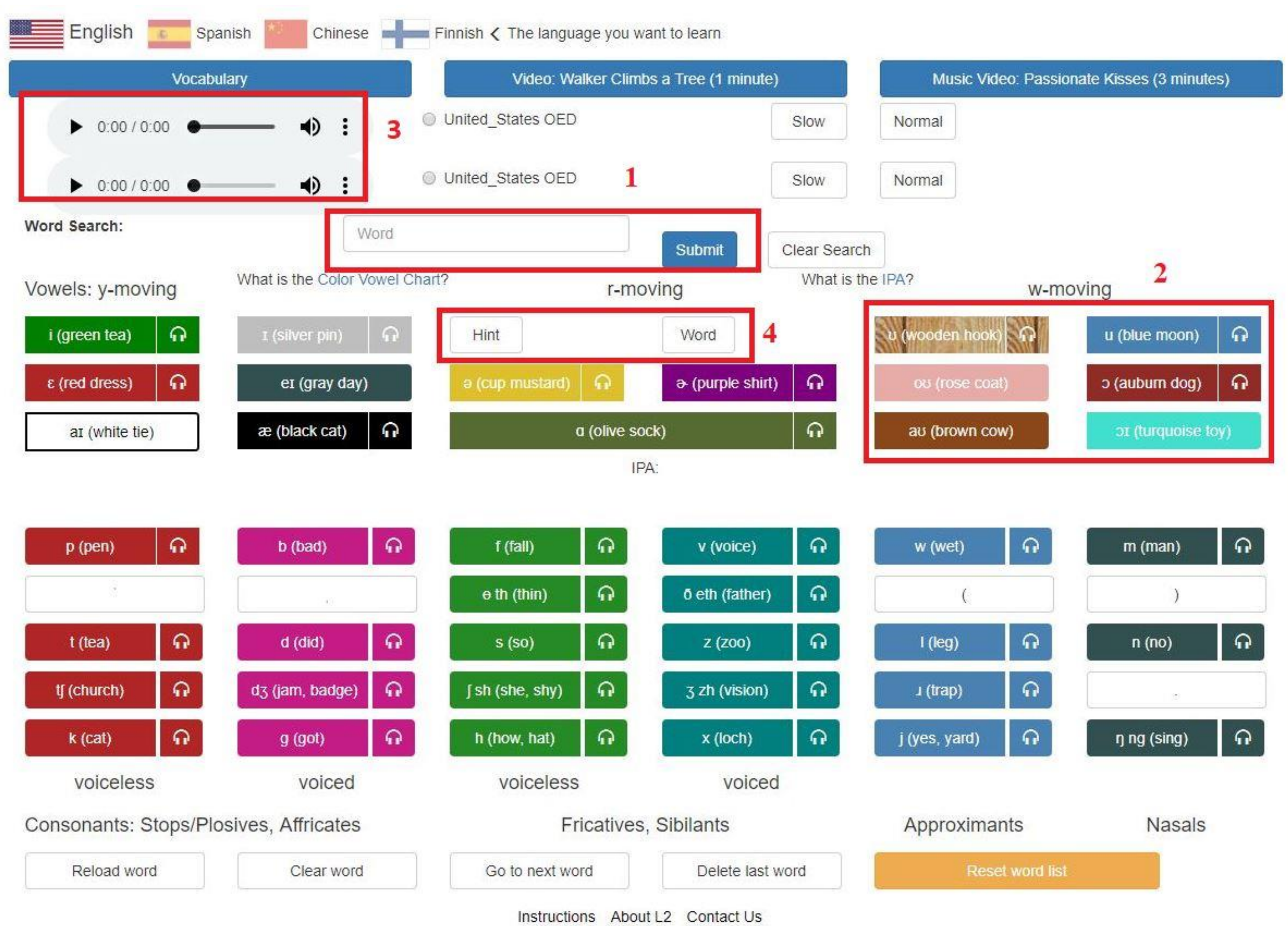


Figure 2. Vocabulary section

Video

In addition to the vocabulary component of the website, LanguageTwo.com also provides short one-minute videos which are divided into smaller parts containing short phrases. The smaller parts are automatically segmented and consist of two-three second videos. The learner can select one short part in the top right corner of the screen and watch them as separate videos. Below the video, the learners can listen to each word that is used in the video, pronounced in isolation, at normal or slower pace (see Figure 3). The website claims that learners can practice their pronunciation of the phoneme they heard and receive immediate feedback, but the immediate feedback feature could not be found on the website at the time of publication (May, 2019). The IPA chart always remains below the exercises so that students can refer to it and listen to individual sounds at any time. Also, detailed instructions are provided in the menu in the top right of the screen if more help is needed. This interface can be seen below in Figure 3.

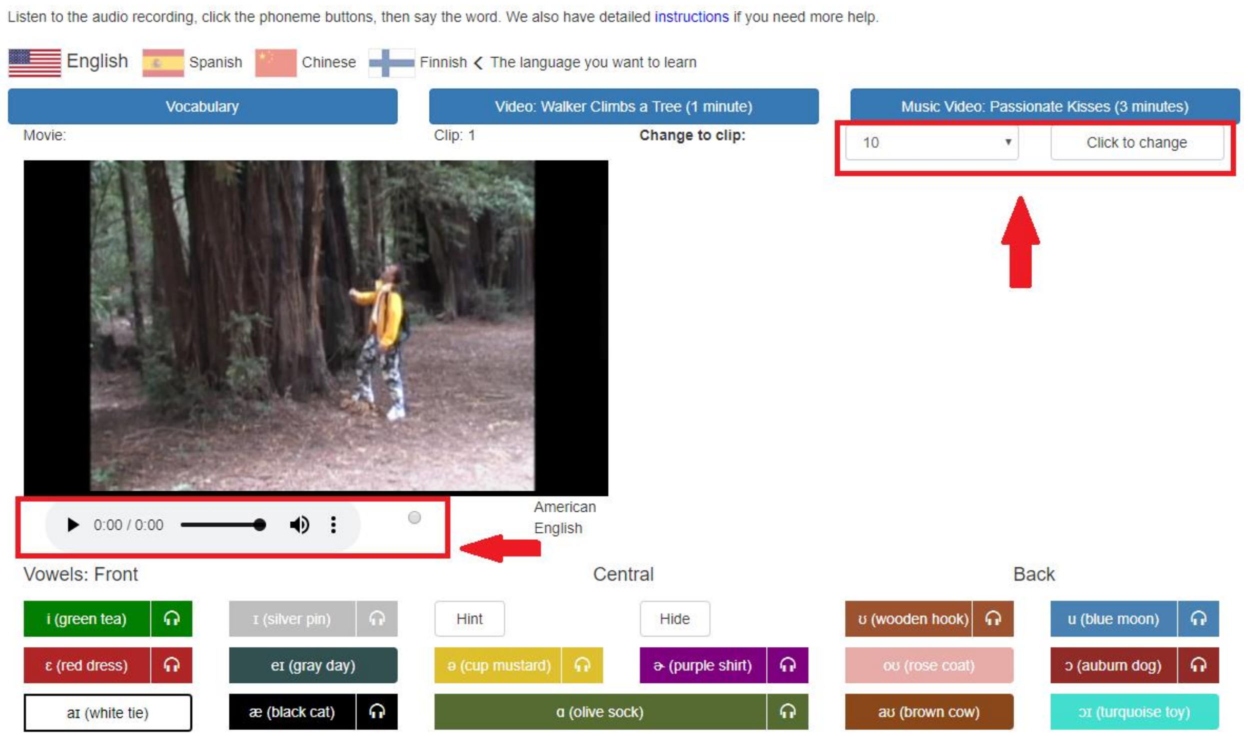


Figure 3. Video section

Music Video

Lastly, learners can listen to and watch a short music video which is broken into 20 shorter chunks which are two - three seconds long each (see Figure 4 below), following the same procedure as in the previous section described above. The website breaks the videos down into phrases, words, and then phonemes. After that, the website suggest that learners should be able to build them back into words, phrases, and stories.

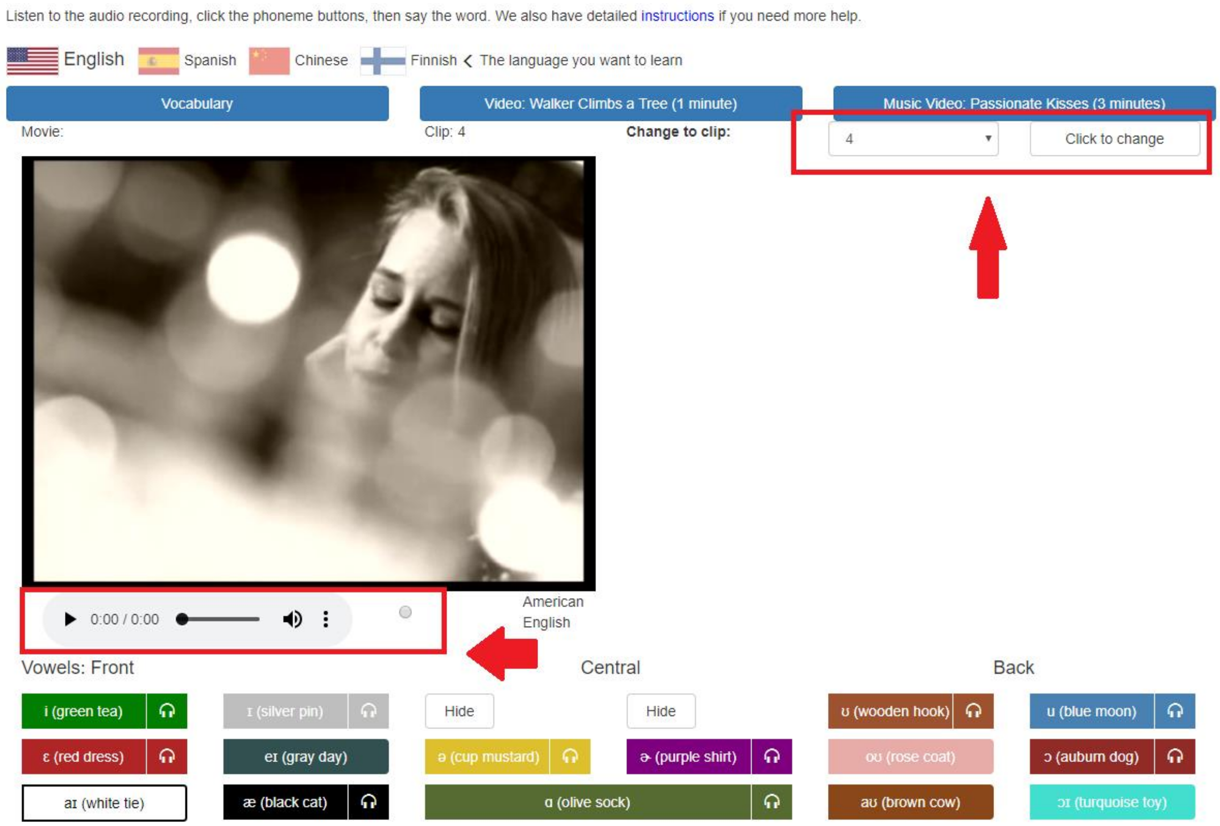


Figure 4. Music video section

After listening to each short phrase within the music video, learners are given the option to take a pronunciation test (see Figure 5). The purpose of this option is to allow learners to test their pronunciation production and receive feedback by Automated Speech Recognition (ASR) system, most likely, to test whether improvement of their perception skills, resulted in improvement of their production skills. It is difficult to find instructions and explanations for this feature on the website, as well as the feature itself. Considering that the website is still developing, this feature may be changed in the future.

Figure 5. Pronunciation test

EVALUATION

As seen through the various screenshots throughout this review, LanguageTwo.com is visually stimulating and draws attention to the IPA chart. The colorful appearance and inclusion of videos may provide an engaging learning environment for the users. In terms of usability, at first glance, the website seems user-friendly and it provides detailed instructions for users who need guidance using the website. The IPA chart is fairly easy to use; nonetheless, users may need to spend some time learning how to use the videos, since that part is a bit confusing. The division of the videos into smaller two-three second chunks and the way it can help the learners may seem difficult to grasp at the beginning, so it would be helpful if this is elaborated upon in the instructions. Occasionally, certain technical problems arose, such as not being able to hear the sound when clicking on a few phonemes. Hopefully these issues will be resolved as soon as the website is fully completed.

One of the important aspects of pronunciation learning is the promotion of learner autonomy (McCrocklin, 2016) and this website encourages autonomous learning as the learners can practice any time and place, at their convenience. The practical aspect of the website, being free and accessible on any device, can also provide support for autonomous, self-paced learning.

Despite this potential for learner autonomy, the website only emphasizes segmental acquisition and there is no information about suprasegmentals. Focusing on segmentals can be useful and effective for beginners who do not have a lot of experience with differences in pronunciation between their first language (L1) and English (Eskenazi, 1999).

However, unlike other software that focuses mostly on production, this website would mostly be useful for speech perception improvement because its main aim is to enhance learners' perception. Badin, Bailly, and Boë (1998) state that second language learners (L2) can be considered phonologically deaf as they may not be able to discriminate between sounds that do not belong to their phonological inventory. Hence, if learners cannot perceive the sounds, they cannot produce them consistently either; thus, if the primary aim of the website is improving perception, then it may serve as a stepping stone towards production improvement.

LanguageTwo.com claims that it can be used to train speech perception and pronunciation production. While Thomson (2011) argues that the amount and quality of L2 input influences L2 production, but we are doubtful that this website can also be useful for production improvement. Even though a feature for pronunciation practice is provided, that feature was not functional when the website was tested. Levis (2007) states that feedback is lacking in Computer-Assisted Pronunciation Training (CAPT) systems as they are usually unable to automatically and accurately diagnose pronunciation. As noted earlier, the website is still under development, and when this feature becomes functional, we hope it will provide accurate, automated feedback to learners. With this feature available, the website would be a suitable resource that might also become useful for production improvement.

For future improvements, we suggest the inclusion of High-Variability Phonetic Training (HVPT) which relies on input produced by multiple contexts and by multiple speakers (Qian, Chukharev-Hudilainen & Levis, 2018) as well as a larger diversity of videos on different topics. As stated by the author, future versions will include data collection so that researchers can study how speech perception affects pronunciation with large data sets of thousands of learners and words.

CONCLUSION

All said, with a few suggested improvements, LanguageTwo.com has the potential to be a successful tool that can be used to aid English as a Second Language (ESL) and English as a Foreign Language (EFL) classrooms around the world. L2 perceptual training is important and necessary, not only because L2 perception is challenging for adult learners, but because it also facilitates oral production (Qian et al., 2018). This website can be used for beginners to introduce them to the sounds in English, but also with more advanced learners who struggle with pronunciation. Finally, the use of this website could facilitate autonomous learning in an engaging way for learners.

REFERENCES

Badin, P., Bailly G., & Boë L. J. (1998). Towards the use of a virtual talking head and of speech mapping tools for pronunciation training. *Proceedings of the ESCA Tutorial and Research Workshop on Speech Technology in Language Learning*, *98*, 167-170.

Eskenazi, M. (1999). Using a computer in foreign language pronunciation training: What advantages? *CALICO Journal*, *16*(3), 447–470.

Levis, J. (2007). Computer technology in teaching and researching pronunciation. *Annual Review of Applied Linguistics*, *27*, 184-202.

McCrocklin, S. M. (2016). Pronunciation learner autonomy: The potential of Automatic Speech Recognition. *System*, *57*, 25–42

Qian, M., Chukharev-Hudilainen, E., & Levis, J. (2018). A system for adaptive high-variability segmental perceptual training: Implementation, effectiveness, transfer. *Language Learning & Technology*, *22*(1), 69–96. doi:10125/44582

Thomson, R. I. (2011). Computer-assisted pronunciation training: Targeting second language vowel perception improves pronunciation. *CALICO Journal*, *28*(3), 744–765.

He, L., & Smith, J. (2019). ImmerseMe [Review]. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Ames, IA, September 2018 (pp. 461-466). Ames, IA: Iowa State University.

TECHNOLOGY REVIEW

ImmerseMe

Long He, Iowa State University
Jordan Smith, Iowa State University

INTRODUCTION

ImmerseMe is an online language-learning platform that offers language instruction in a variety of virtual reality–based settings. The use of virtual reality allows for more authentic representations of real-life interactions learners may find themselves in if they travel to a region where the target language is spoken. Learners select a setting and a lesson, and then virtually interact with prerecorded interlocutors. Novice learners can also participate in dictation exercises in which they repeat the words spoken by their virtual interlocutor. The speech that learners produce is recorded by the computer's microphone, transcribed automatically, and evaluated by the application. The conversation adapts based on the available responses users choose as they work through the lessons.

In this review, we provide a brief overview of the tool and then offer an evaluation of its affordances. Since the *ImmerseMe* application that makes use of VR headsets is not slated for release until late 2019, this evaluation is based on ImmerseMe's 2018 Google Chrome desktop application.

OVERVIEW

Currently, *ImmerseMe* offers learning materials for nine languages: German, Spanish, French, English, Japanese, Chinese, Italian, Greek, and Indonesian. After a learner has selected their[i] desired language, they are invited to select a lesson based on a given communicative situation. Figure 1 shows some of the possible lessons German learners can select, including buying coffee in a café or checking out at a chocolate shop.

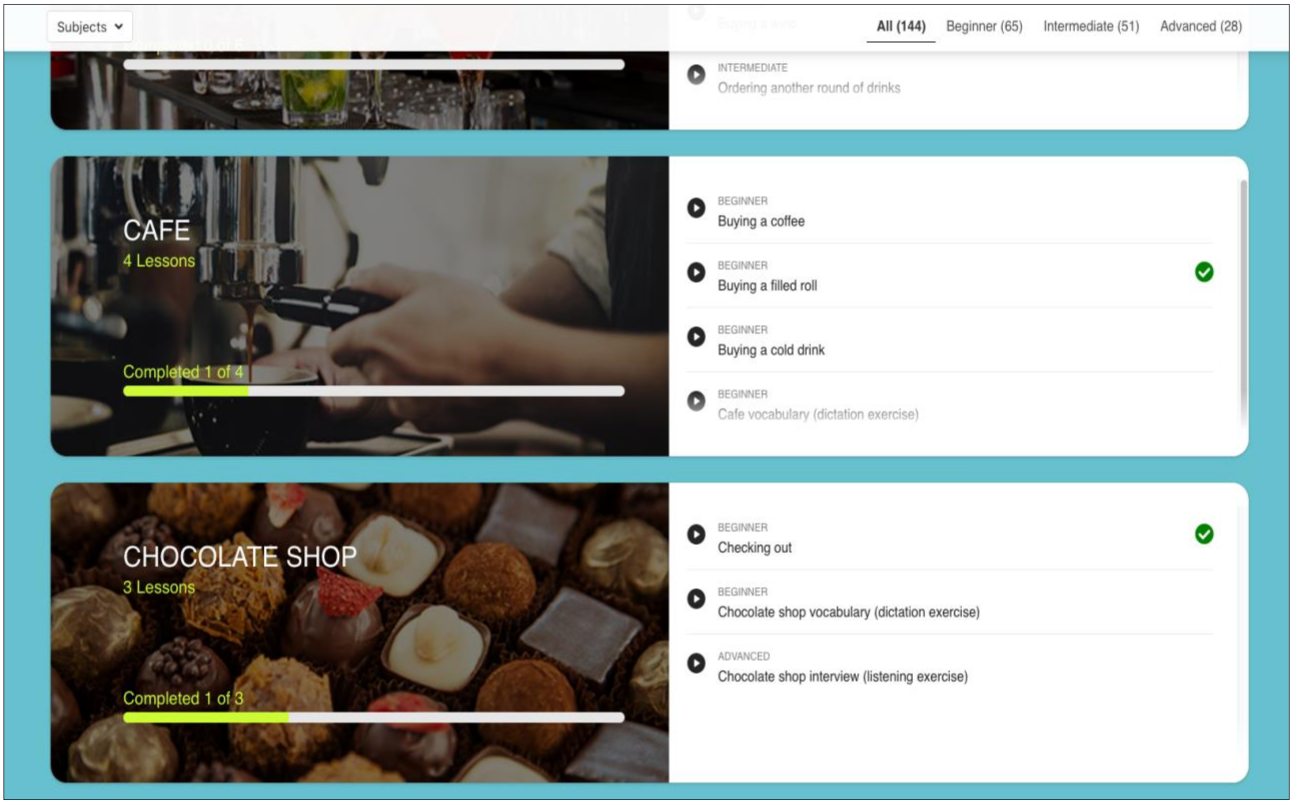


Figure 1. German-language lessons from ImmerseMe.

Once the learner selects their desired situation, they are then presented with a transcription of the interaction that will take place as part of the lesson (Figure 2). The transcription appears in both the target language and in English, allowing learners to preview the vocabulary they will need to know in order to successfully communicate in the context.

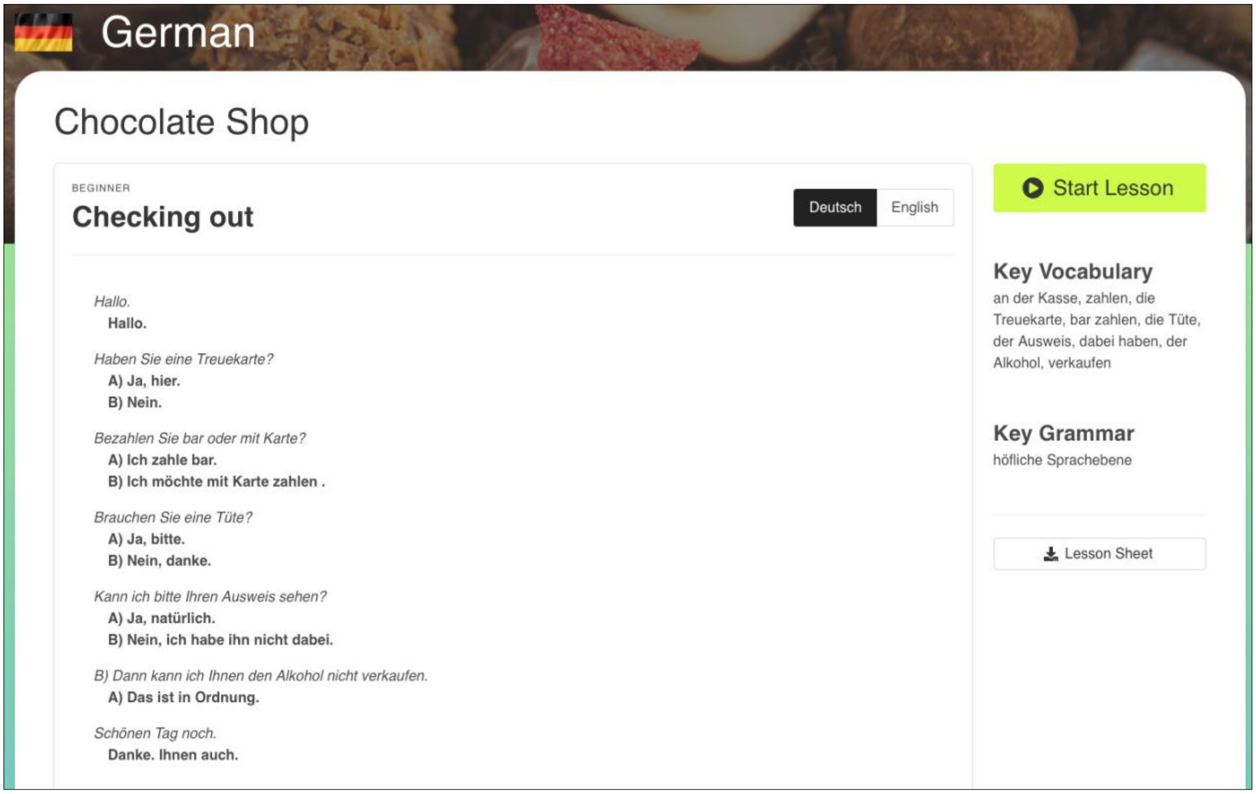


Figure 2. Transcript from one German lesson.

After the learner starts the lesson, they watch a video intended to represent a real-life communicative event (Figure 3). A transcript of the interlocutor's speech appears at the top of the screen. Once the interlocutor has completed speaking, the learner can select from one or more possible responses, transcribed in the middle of the screen. Clicking on the green microphone button at the bottom of the screen activates the learner's microphone. The learner's speech is recorded, transcribed in the field at the bottom of the screen, and evaluated. If the speech does not satisfactorily match the expected pronunciation, the learner must repeat the response.

Figure 3. Virtual interaction with a native German speaker on ImmerseMe.

All videos have been filmed using a 360-degree camera and are shot from the learner's point of view. Learners using a standard computer can use their mouse to change the view to see what is around them. Learners using virtual reality goggles benefit from an even more realistic representation of the situation, as simply shifting the direction of their gaze will change their perspective of the situation.

EVALUATION

Chapelle (2001) argues that an evaluation of a CALL task—and, by extension, tool—"cannot be a categorical decision about effectiveness" but should instead be "an argument indicating in what ways [it] is appropriate for particular learners at a given time" (p. 53). To build an argument for evaluating a CALL task, Chapelle outlines six criteria: *language learning potential, learner fit, meaning focus, authenticity, positive impact*, and *practicality*. In this section, we evaluate the website *ImmerseMe*, focusing our evaluation on what we feel are the most salient criteria from Chapelle's framework.

Language learning potential and meaning focus

Chapelle (2001) refers to *language learning potential* as "the extent to which the task promotes beneficial focus on form" (p. 55). "Focus on Form" (Long, 1991) emphasizes the need for learners to pay attention to various aspects of linguistic form while simultaneously engaging in meaning-focused communication activities. The tasks included in *ImmerseMe* lessons are intended to simulate real-life interactions in virtual representations of authentic settings. However, providing learners with preselected options to read from as they interact with the virtual interlocutor undermines the meaning-focused nature of the tasks. While some learners may concentrate on the meaning of the interactions, others will likely focus only on the pronunciation of the sentences they are given without paying attention to meaning. It is therefore uncertain whether an adequate focus on meaning could be achieved in the tasks.

Levis (2007) points out that "[computer-assisted pronunciation teaching] systems often suffer from difficulties in giving learners adequate, accurate feedback and an inability to provide accurate and automatic diagnosis of pronunciation errors" (p. 185). Corrective feedback, one of the main techniques to accomplish focus on form (Nava & Pedrazzini, 2018), is also a feature that needs to be more organically integrated into the *ImmerseMe* application. Though the application offers a live speech-to-text preview, which is claimed to provide students with immediate feedback about their pronunciation, the voice-recognition technology does not seem to provide accurate transcription. For example, we noticed that the final word of an utterance was sometimes left out of the transcription, and that even if the pronunciation of an utterance was intentionally incorrect, the program leniently considered it as passable and proceeded to the next task, which suggests a substantial weakness in its requirement for users to "pronounce the correct answer perfectly" (as quoted in Lucente, 2018, p. 4) before advancing to the next stage. Some words in the transcription appeared in a red color, suggesting an error or other problem with pronunciation, though red-colored words did not always seem to have an effect on a learner's ability to successfully complete a lesson. As a result, the feedback offered by the program is lacking, and it is not clear how the pronunciation is evaluated.

Authenticity

Chapelle (2001) defines *authenticity* as "the degree of correspondence between an L2 learning task and tasks that the learner is likely to encounter outside the classroom" (p. 56). The design of *ImmerseMe*'s communicative events aims to put learners in an immersive environment where they can have guided interaction with native speakers. The tasks are situation specific and videos were recorded at a normal speech rate. In these ways *ImmerseMe* allows learners to feel as if they are in an authentic speaking situation (see Bajorek, 2018, for a user comment to this effect). However, the authenticity of *ImmerseMe*'s tasks suffers in important ways. For example, as noted above, the tasks do not necessarily require users to comprehend the speech they hear, as they can choose to read a transcription and/or a translation of their interlocutor's speech. Because learners are given a fixed set of responses to choose from, the tool does not allow learners to practice producing authentic responses. Instead, it may encourage learners to simply read from the screen, which can have value for improving pronunciation but not for improving communicative competence. Learners, particularly intermediate- or high-level learners, need opportunities to mobilize their linguistic resources to negotiate meaning because meaning negotiation can push learners to produce more target-like utterances (Long, 1991).

Positive impact and practicality

According to Chapelle (2001), tasks "should help learners to gain pragmatic abilities that will serve in communication beyond the classroom" (p. 57). In other words, learners could transfer what they have learned from the task to other communication scenarios. Practicing the language in a virtual scenario that mimics real-life interactions prepares learners for similar situations they will encounter in the real world. *ImmerseMe* also offers self-conscious learners who are hesitant to engage in conversations with native speakers a safe space to use the language in a virtual environment until they have built up their confidence. However, these positive effects are not available to everyone. Since the videos are hosted on YouTube, students in China or other areas where the site is blocked do not have access to them, limiting the application's reach.

CONCLUSION

Virtual-reality technology extends learners' access to authentic language-learning experiences. *ImmerseMe* creates a contextualized environment where learners can enhance their language skills. In this review, we have offered an overview of the tool as well as an evaluation of some of its affordances. As Chapelle (2001) points out, empirical analysis is also an indispensable part in constructing an evaluation argument. As such, more empirical research on how virtual reality–enabled language-learning tools impact learning outcomes is needed.

REFERENCES

Bajorek, J. (2018). Virtual Reality Technology ImmerseMe Software: Speaking, Interactivity, & Time [PDF]. Retrieved from https://www.academia.edu/36768999/Virtual_Reality_Technology_ImmerseMe_Software_Speaking_Interactivity_and_Time._Bajorek_2018_CALICO_2018.

Chapelle, C. A. (2001). *Computer applications in second language acquisition*. Cambridge: Cambridge University Press.

Levis, J. (2007). Computer technology in teaching and researching pronunciation. *Annual Review of Applied Linguistics, 27*, 184–202.

Long, M. (1991). Focus on form: A design feature in language teaching methodology. In K.D. Bot, R. Ginsberg, & C. Kramsch (Eds.), *Foreign language research in cross-cultural perspective* (pp. 39–52). Amsterdam: John Benjamins.

Lucente, A. (2018, August 12). VR program brings languages to life. *The Age*. PDF retrieved from https://immerseme.co/

Nava, A., & Pedrazzini, L. (2018). *Second language acquisition in action: Principles from practice*. New York, NY: Bloomsbury Academic.

[i] In this paper, we use *they* as a gender-neutral singular pronoun.

Neiriz, R. (2019). The Phonetics 3D [Review]. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Ames, IA, September 2018 (pp. 467-472). Ames, IA: Iowa State University.

TECHNOLOGY REVIEW

The Phonetics 3D

Reza Neiriz, Iowa State University

INTRODUCTION

Developed by Fuminori Homma and Yasushi Okumura and edited by Masashi Negishi, a professor in Graduate School of International Studies and Institute of Global Studies at Tokyo University of Foreign Studies, *The Phonetics 3D* is a mobile app designed to teach pronunciation to English language learners (ELL). The second version of this app, published in 2018, is available for only iOS users at $7.99. Below, I will first explain how the app works, and then I will present a critical review of the app.

How it works

The first run of the app launches a step-by-step tutorial of how to use it, which can be skipped and accessed later. The homepage has a navigation bar on the top of the screen dividing the content of the app into basic and advanced levels. Under the Basic tab, there is a list of phonemes written in IPA (Figure 1). These phonemes are grouped into stops, fricatives, nasals, laterals, semi-vowels, short vowels, long vowels, and diphthongs, and each group has a short description of its characteristics. The advanced tab contains instruction on phonological features of different phonemes in different environments (Figure 1). It is comprised of consonant clusters, devoicing, unreleased stops, nasal release, lateral release, assimilation, elision, reduction, and aspiration. Each section has a short description of the phonological feature and a list of buttons each representing the target phoneme(s).

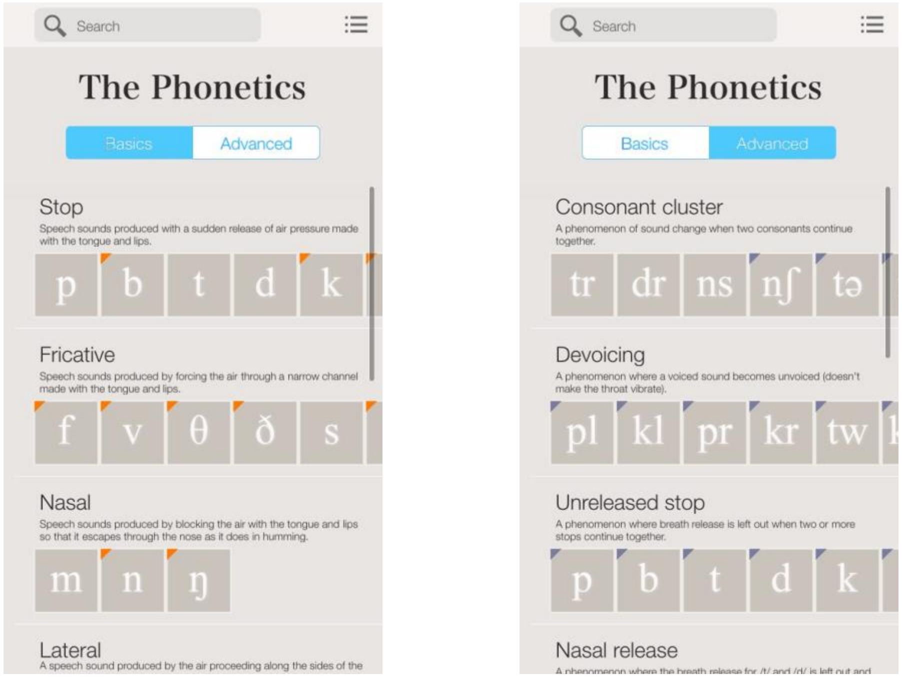


Figure 1. Basic (left) and advanced (right) menus.

Selecting each of the phonemes under the Basic tab takes the user to a new page which has an augmented reality head (Figure 2). This is simply a 3D image of a head visible from the tip of the nose to the throat. It is transparent and unicolor, but different tones of grey highlight the articulators. Users can swipe right or left to have a frontal or side view of this head. To the bottom-left of the head, there are two buttons representing the names of two native speakers whose voices could be used for pronunciation, i.e. Mike and Judy (Matthew K. Miller and Victoria Pate). Below the 3D image box, there are steps which explain articulation mechanisms accompanied by still image icons of side-view cross-section of a head in which blue areas represent the shape of the vocal tract at the respective stage of articulation. The 3D head on top of the screen reinforces this by showing an animated version of this explanation. When the user taps on the 3D image, the entire articulation is illustrated in animation, and tapping on each step illustrates the articulation mechanism of only that step. The animated head uses a white color for the airflow in voiceless phonemes and a blue color for their voiced counterparts. At the bottom of this page, there are two more tabs, one of which has example words containing the target phoneme both spelled and transcribed using IPA. Tapping on these words plays the word along with the 3D illustration of articulation through the transparent head. The last tab is similar to the second tab, but it contains words with phonemes that are similar to the target phoneme in terms of the place of articulation. For instance, for /m/, the software lists words with /p/ and /b/ as they are bilabial like /m/.

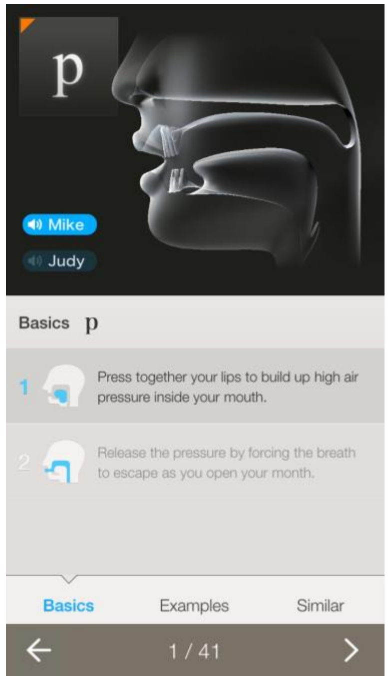


Figure 2. Instructions and modeling page of the Basic tab items.

The instruction page of the phonetic features under the Advanced tab is extremely similar to that of the phonemes (Figure 3). There are, however, two differences. First, there is an added written explanation of the phonological feature, such as how /t/ becomes more like /ʧ/ in *try*. Second, it contains only the examples tab, which lists words containing this feature. The user can tap on each word to both listen to and see the animated articulation of the word. Users can also choose to play the word slowly, and the app resynthesizes the original pronunciation sound file to produce a low-speed playback.

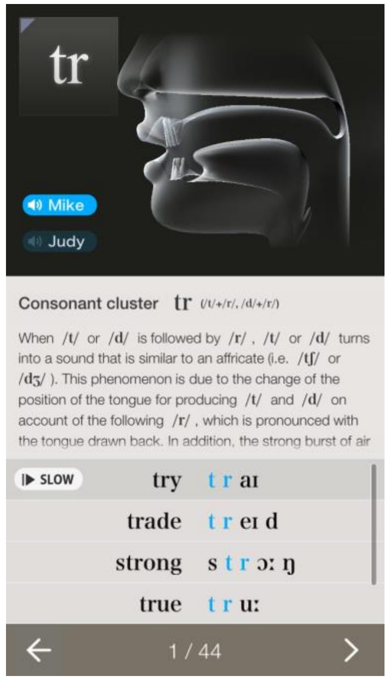


Figure 3. Instructions and modeling page of the Advanced tab items.

One last feature of the app is its search function. Users can type in any word and see both the IPA transcription and use the audio-visual representation of it through the head and either the voices of the speakers of the app or Siri if the word is not already recorded in the app (Figure 4).

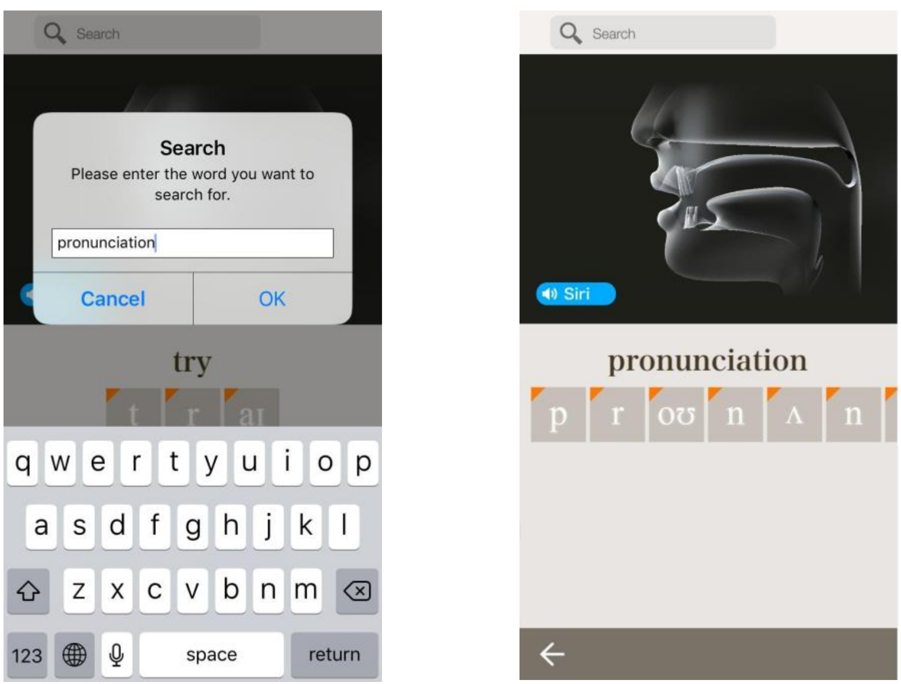


Figure 4. Word search page (left) and 3D articulation simulation of the search result (right).

CRITICAL REVIEW

According to The Phonetics 3D website, this app is "The most beautiful learning tool designed for those who are learning the pronunciations of the English language." While the aesthetics are admirably well-designed, the audience mentioned by the developers is too broad. The use of IPA phonetic symbols along with reference to articulators and articulatory mechanisms can certainly render this app less than useful for beginner and intermediate language learners, and even for advanced learners of English without formal training in phonetics and articulatory anatomy and mechanisms. Perhaps these features make this app more suitable for both native and non-native English speaking teachers. For the former, while most of the articulation involves tacit knowledge, this app can help them turn this knowledge into a procedural form so that they can transfer it to their students. As to non-native teachers, they can refine their pronunciation, both at declarative (knowing how to pronounce) and procedural (being able to actually perform the pronunciation) levels to provide a better role model and instruction of pronunciation for their students.

The goal of a 3D-animated representation of articulatory mechanisms through a transparent virtual head might have been to support auditory instruction. However, the literature is not conclusive in this regard. Similar virtual head representations to teach pronunciation have been used before such as ARTUR (Engwall & Balter, 2007), Baldi (Massaro & Light, 2003), and MASSY (Fagel & Madany, 2008) with inconclusive results. Massaro and Light (2003) compared the effect of using Baldi's face with that of another virtual head whose vocal tract and articulators were visible in improving Japanese speaker's pronunciation of /r/ and /l/. The results showed no difference. In another study, Massaro, Bigler, Chen, Perlman, and Oui (2008) used a virtual head with a visible vocal tract to teach Arabic consonants. The difference between virtual head users and the group presented with only auditory input was negligible. The results, however, have not always been negative. Massaro and Light (2004) found improvements in hearing-impaired American children through audio-visual training of consonant clusters, fricative-affricate, and voicing distinctions. Fagel and Madany (2008) also found improvements in children's lisping after using an augmented reality head to teach the articulation of /s/ and /z/. In short, Engwall (2012) argues that using audio-visual articulation training is not necessarily conducive to improvement in pronunciation if it is not accompanied by feedback. All of these studies show that using *The Phonetics 3D* might only be useful if it is used in combination with feedback from teachers who are familiar with phonetics and phonology. In other words, it can be a useful teaching aid in pronunciation classes where teachers can focus on only the problematic areas while all students spend time on the targeted practice of pronunciation instead of following classroom-wide drills which might not benefit everyone equally.

While *The Phonetics 3D* is a promising tool in aiding teaching pronunciation, it is lacking a major component of pronunciation, the suprasegmental features. Although this app has included phonological features, such as co-articulation, students interested in improving their pronunciation to approximate native-like speech could have benefitted from the inclusion of suprasegmental features in the app. These features have been shown to be strong indicators of accent. For instance, Van Els and de Bot (1987) found that even with low-pass filtered speech (retaining suprasegmental features) and monotonized speech (retaining segmental features), the participants could detect the foreign accent in speakers. Therefore, the inclusion of suprasegmental training material in *The Phonetics 3D* would be a welcome addition.

CONCLUSION

The Phonetics 3D comes with useful features, such as transparent and animated representation of vocal tract and the use of different colors for voiced and devoiced features, and it can be a useful pronunciation learning tool when accompanied by instruction and feedback. Yet, an addition of suprasegmental features to this app can better respond to the needs of users who wish to achieve a more native-like pronunciation.

REFERENCES

Engwall, O. (2012). Analysis of and feedback on phonetic features in pronunciation training with a virtual teacher. *Computer Assisted Language Learning, 25*(1), 37-64.

Engwall, O., & Balter, O. (2007). Pronunciation feedback from real and virtual language teachers. *Computer Assisted Language Learning, 20*, 235-262.

Fagel, S., & Madany, K. (2008). A 3-D virtual head as a tool for speech therapy for children. In International Speech Communication Association (Ed.), *9th annual conference of the International Speech Communication Association 2008 (Interspeech 2008)* (pp. 2643-2646). Red Hook, NY: Curran Associates, Inc.

Massaro, D. W., & Light, J. (2003). Read my tongue movements: Bimodal learning to perceive and produce non-native speech /r/ and /l/. *Eurospeech* (pp. 2249-2252). Brisbane, Australia: European Speech Communication Association (ESCA).

Massaro, D., & Light, J. (2004). Using visible speech for training perception and production of speech for hard of hearing individuals. *Journal of Speech, Language, and Hearing Research, 47*, 304-320.

Massaro, D., Bigler, S., Chen, T., Perlman, M., & Oui, S. (2008). Pronunciation training: The role of eye and ear. *Interspeech* (pp. 2623-2626). Geneva, Switzerland: International Speech Communication Association (ISCA).

van Els, T., & de Bot, K. (1987). The role of intonation in a foreign accent. *The Modern Langauge Journal, 71*(2), 174-155

Ozkul, A. (2019). Accent Perfect: American English Pronunciation App [Review]. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Ames, IA, September 2018 (pp. 473-482). Ames, IA: Iowa State University.

TECHNOLOGY REVIEW

Accent Perfect: American English Pronunciation App

Altay Ozkul, Iowa State University

INTRODUCTION

Second language learning can be fostered when language learners pay attention and notice the target language input and structures and understand the importance of what they notice (Schmidt, 2012). Computers may promote autonomous learning through individualized instruction, practice through listening and repetition exercises and automatic visual support (Levis, 2007). There have been numerous computer and smartphone apps developed as additional resources for language learners; these applications seek to aid users in becoming self-supporting learners and practice their language skills outside of traditional classrooms.

In this context, *Accent Perfect: American English Pronunciation App* is an app for iOS and Android smartphones that was developed by MNS C DEV LLC in 2016 (AppBrain). The app can be downloaded from iTunes Store or Google Play for $19.99. The developer claims that the app seeks to help ESL learners develop reading, listening, speaking and intonation skills (MNSC). The app seeks to achieve this goal by having its users “recognize common words that share the same sound with the help of [its] guided audio recordings and practice sessions” (MNSC).

Overview

The app consists of four sections: (1) sounds, (2) minimal pairs, (3) intonation and (4) games. When the user launches the app for the first time, they are asked to specify their native language. The app only gives the user seven language options to choose from: German, French, Chinese, Spanish, Hindi, Korean, Italian and an “English or Other” option for all other languages (Figure 1).

Figure 1. Language selection menu.

Once the user has indicated their native language, they are directly taken to the main interface without any training on how to use or navigate the app. However, in the main menu, the user can open "Instructions," which gives them a brief overview on how they can utilize the app's functions (Figure 2).

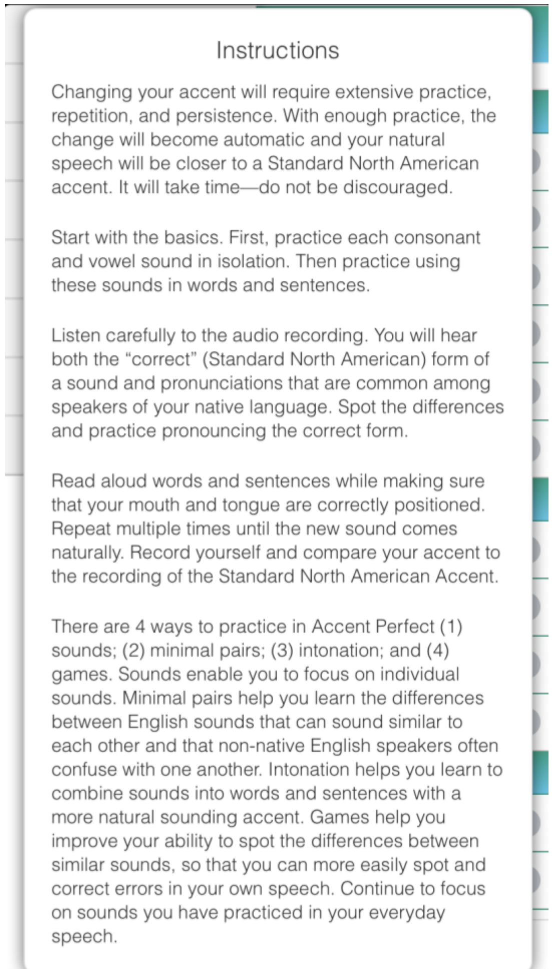


Figure 2. Instructions on how to use the app properly.

After reading the instructions, the user selects between "Arpabet" (orthographies) and "International Phonetic Alphabet" (IPA) symbols to represent consonants and vowels (Figure 3). "Phonetic system" is also among the options that the user can select but selecting "Phonetic system" and "Arphabet" result in the same representations of sounds.

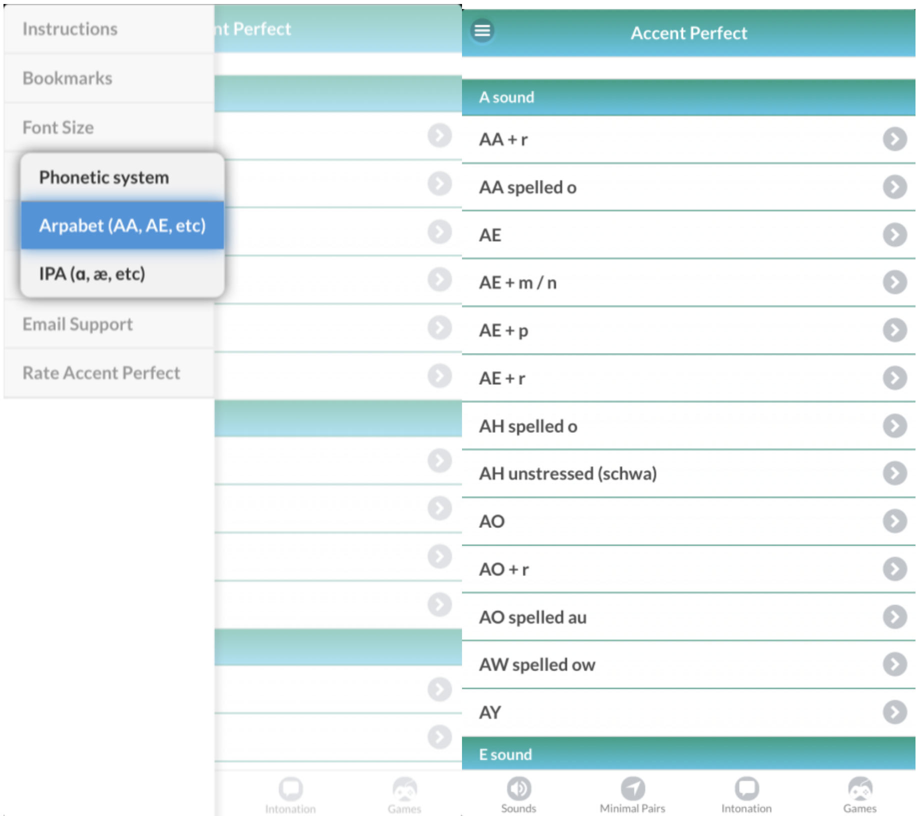


Figure 3. Phonetic system selection and the main menu.

As can be seen in Figure 3, the user can navigate by touching "Sounds," "Minimal Pairs" or "Intonation" categories at the bottom of the screen. Each category consists of a short audio lecture about the item in question and word- or sentence-level repetition exercises. It is important to note that the user gets to choose from two different voices in the listen-and-repeat exercises, though, both voices are women (Figure 4.)

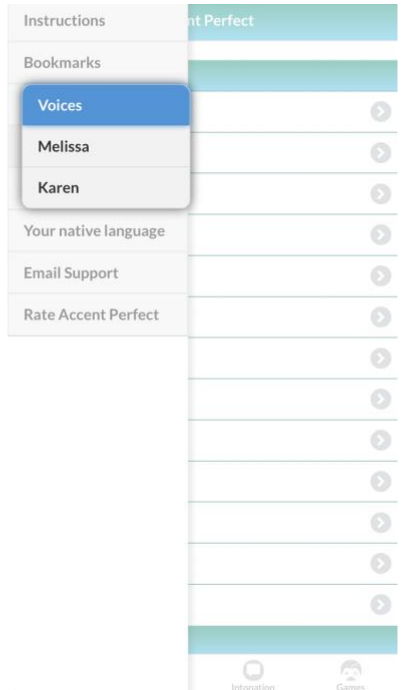


Figure 4. Speaker selection menu.

Sounds

In "Sounds," the user can listen to lectures and complete word- and sentence-level repetition exercises on certain sounds such as consonants, pronunciation of simple past *-ed* that follows various consonants (such as /d/ or /t/), /s/ consonant in final position, and "silent sounds" (such as /n/, /p/, /d/, or /k/). Lectures consist of instructions about why a particular sound may be difficult to produce for some non-native speakers of English and how production of that sound can be improved. Exercises include listening to the correct pronunciation of the sound. The user also gets to record themselves and play the recording to compare their pronunciation to that of the speaker the app provides (Figure 5). It should be noted that the app does not provide any feedback to its users on their performances.

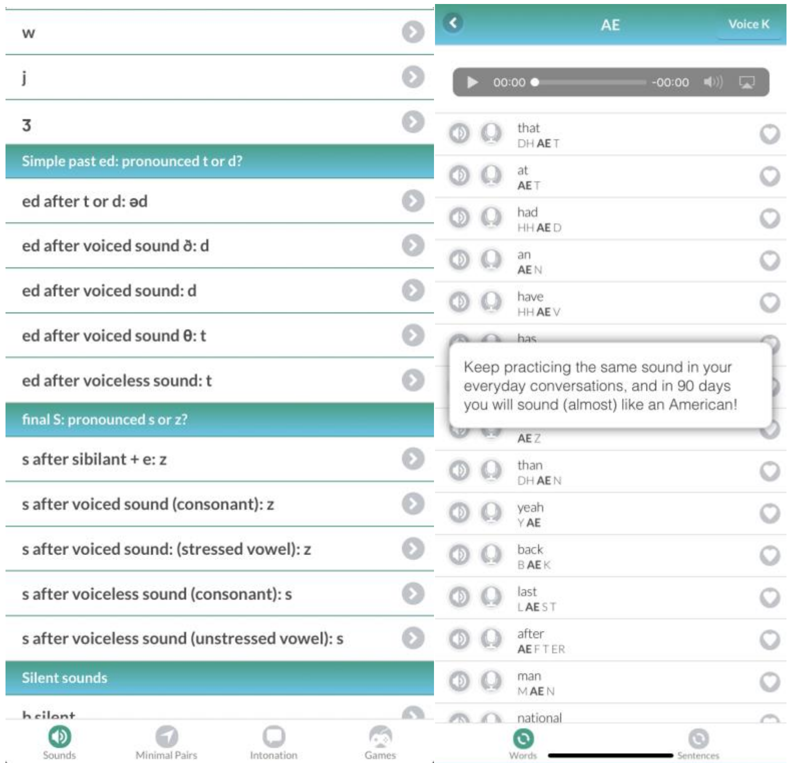


Figure 5. "Sounds" section and exercises.

Minimal pairs

The user can find lectures and repetition exercises on minimal pairs of consonants (such as /s/-/ʃ / or /f/-/p/) and vowels. There is also a section for homophones (such as *bare-bear*, *tear-tier* and *soar-sore*) where the user can listen to similar repetition exercises and practice their pronunciation by recording themselves/ (Figure 6).

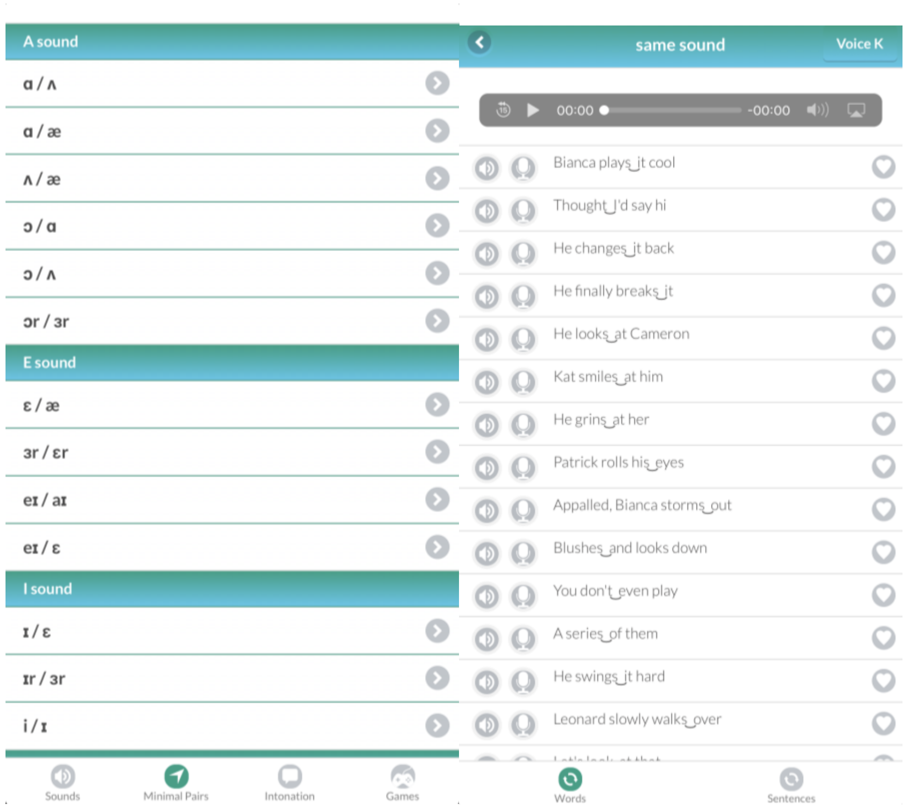


Figure 6. "Minimal Pairs" section and sentence-level exercises.

Intonation

This component includes four main categories of exercises, namely, consonant to vowel, similar consonant, same consonant, vowel to vowel: w and vowel to vowel: y. The user can practice intonation of these elements of connected speech through sentence repetitions; however, the app does not provide a lecture that the user can listen to in this component (Figure 7).

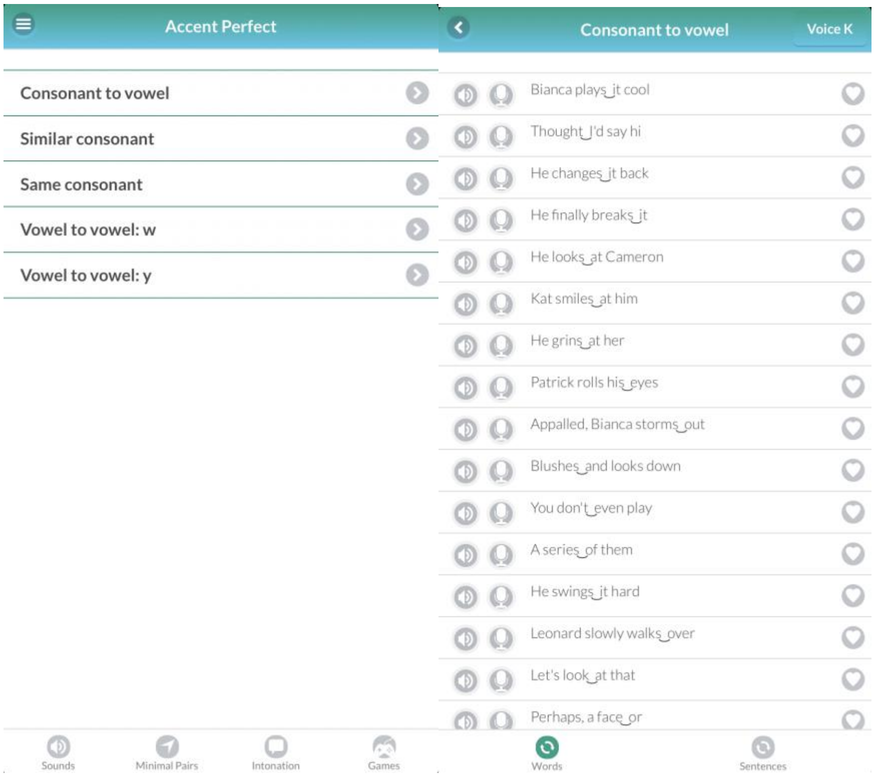


Figure 7. "Intonation" section and exercises.

Games

Games consists of one short where the user listens to the pronunciation of a single word or a keyword uttered in a sentence and is asked to identify the word that they heard. The keywords are based on what the user practices in the four main sections of the app. There is a progression; the user advances if they complete a “level” consisting of fifteen questions without exceeding three wrong answers. However, there are no rewards that the app offers for completing levels, and the levels do not seem to get harder as the user makes progress (Figure 8).

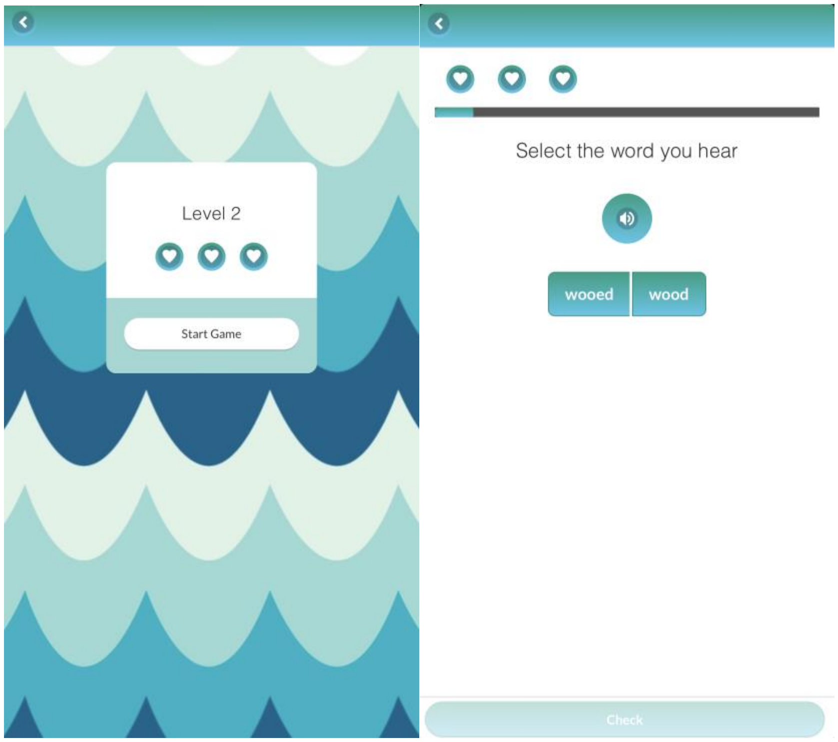


Figure 8. "Games" section of the app.

EVALUATION

The various word- and sentence-level drilling exercises that *Accent Perfect* consists of can be beneficial for learners who are used to learning and improving pronunciation through repetition. Learners may also find two different representation of sounds and lectures with brief instructions on how to produce certain sounds useful. However, what can immediately be noticed in *Accent Perfect* is the lack of variety in the exercise types that it provides. The app only includes listening and repetition exercises but lacks automatic visual support such as graphical displays of the speaker's vocal tract or face that Levis (2007) suggests. It also does not provide any sort of feedback to its users, expecting them to engage in self-evaluation by repeatedly listening to the correct form and their own productive, and try to replicate the correct form without any guidelines. Saito (2012) suggests that pronunciation exercises should not only be in the form of drilling exercises that are strictly controlled; instead, exercises should promote communicative practice to teach learners meaningful, spontaneous speech abilities. *Accent Perfect* is clearly far from achieving what Saito (2012) puts forth as it seems to fail to implement communicative exercises that would foster both pronunciation teaching and speaking ability. Implementation of automated visual and textual feedback through speech recognition would help learners understand their errors and correct themselves better instead of trying to improve their pronunciation on their own (Levis, 2007).

It is also worth mentioning that the app utilizes Arpabet and IPA symbols to represent sounds. Erdener and Burnham (2005) noted that English has an opaque orthography in which one sound may be shown by different orthographic representations. The use of Arpabet can become an issue for many speakers with a native language that has a transparent orthography as they would transfer their L1 knowledge into their L2, which would lead to interference. The app attempts to cope with this issue by including the IPA symbols and by providing lectures and drilling exercises with words that have the same vowel but spelled differently.

The app attempts to employ game-based learning but lacks depth in its implementation. Gros (2007) notes that game-based learning should foster environments in which "skills and attitudes play an important role" (p. 26). What the game in the app offers is essentially more listening and drilling exercises which the user has done (or should do) numerous times before playing the game. There is a need of more exercise types based on game-based learning theories to reduce such repetitiveness and offer varied, meaningful activities that the user can benefit from. If the developers intend is to give the users an opportunity to practice their perception of American English in a simple game, the "Games" section can be renamed to more accurately reflect the nature of the activity that users experience.

The app also lacks authentic tasks representing the pronunciation features found in real-life communication. Chapelle (2001) defines "authenticity" as the connection between computer-assisted language learning (CALL) activities and communication skills that are used outside of the classroom. Therefore, the exercises could be representative of the pronunciation features one can experience in daily life. *Accent Perfect* provides only repetition exercises at word- and sentence-level, failing to deliver meaning- and content-based (Chapelle, 2001) pronunciation exercises to users.

CONCLUSION

Despite its weaknesses in terms of speech recognition, visual feedback and varied communicative exercises, *Accent Perfect* can still be beneficial for a language learner who would like to be exposed to the features of American English outside of classrooms through short lectures, listening activities, and repetition exercises on minimal pairs and intonation of consonants and vowels. The developers might consider inclusion of meaning- and content-based exercises and a progression system in the "Games" section to foster game-based learning. This app will be more beneficial for ESL learners if these issues are addressed based on previous research and theory in the field of pronunciation.

REFERENCES

AppBrain. (nd). Retrieved from https://www.appbrain.com/dev/MNSC/

Chapelle, C. (2001). *Computer applications in second language acquisition.* New York: Cambridge University Press.

Erdener, V. D., & Burnham, D. K. (2005). The role of audiovisual speech and orthographic information in nonnative speech production. *Language Learning, 55*(2), 191-228.

Gros, B. (2007). Digital games in education: The design of games-based learning environments. *Journal of Research on Technology in Education, 40*(1), 23-38.

Levis, J. M. (2007). Computer technology in teaching and researching pronunciation. *Annual Review of Applied Linguistics, 27,* 184-202.

MNSC. (2016). Accent Perfect: American English Pronunciation App. Retrieved from https://itunes.apple.com/us/app/accent-perfect-american-english-pronunciation-app/id959744135?mt=8

Saito, K. (2012). Effects of instruction on L2 pronunciation development: A synthesis of 15 quasi-experimental intervention studies. *TESOL Quarterly*, *46*(4), 842-854.

Schmidt, R. (2012). Attention, awareness, and individual differences in language learning. In W. M. Chan, K. N. Chin, S. K. Bhatt, & I. Walker (Eds.), *Perspectives on individual characteristics and foreign language education* (vol. 6, pp. 27-51). Boston: De Gruyter Mouton.

Silva dos Santos, L. (2019). RachelsEnglish.com [Review]. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Ames, IA, September 2018 (pp. 483-489). Ames, IA: Iowa State University.

TECHNOLOGY REVIEW

RachelsEnglish.com

Liberato Silva dos Santos, Iowa State University

INTRODUCTION

Pronunciation training has increasingly been recognized for the important role it plays in making oral communication intelligible and meaningful for both listeners and speakers (Levis, 2017). This recognition is in part responsible for the growing number of online resources such as websites, blogs, and online courses that offer some kind of pronunciation support. Some of these platforms only have lists with learning tips while others provide audio and video tutorials and some can interact with learners. A closer look at these resources can help us see what they can and cannot do, and this is our aim as we look at one very popular online resource known as Rachel's English.

Overview

Rachel's English is a combined website, podcast, and online course on speaking and pronunciation instruction with a focus on American English. It provides over 400 free instructional videos that cover topics such as how to pronounce each sound in the IPA chart and how to work on rhythm, intonation, linking, and stress when uttering words, phrases, sentences, and larger units of speech. The intended audience includes "non-native speakers wanting to work on their spoken English for job advancement or to pass a test; teachers, native and non-native alike, who teach ESL; Americans who have suffered a stroke and need resources to re-learn the movements of the mouth when speaking English" (Rachel's English, 2018). Users can subscribe to a newsletter and be notified of updates and additions to website content, such as new videos. Rachel's English YouTube channel boasted 1.9 million subscribers and over 90 million views as of April 2019. The accompanying Facebook page had approximately 329,000 followers and 319,000 likes around the same time period. The website is a hub from which users can access several different resources (see Figure 1). The top of the website's homepage has links with labels such as Videos, Courses, Podcast and (external) Resources, each of which will be briefly discussed in this review.

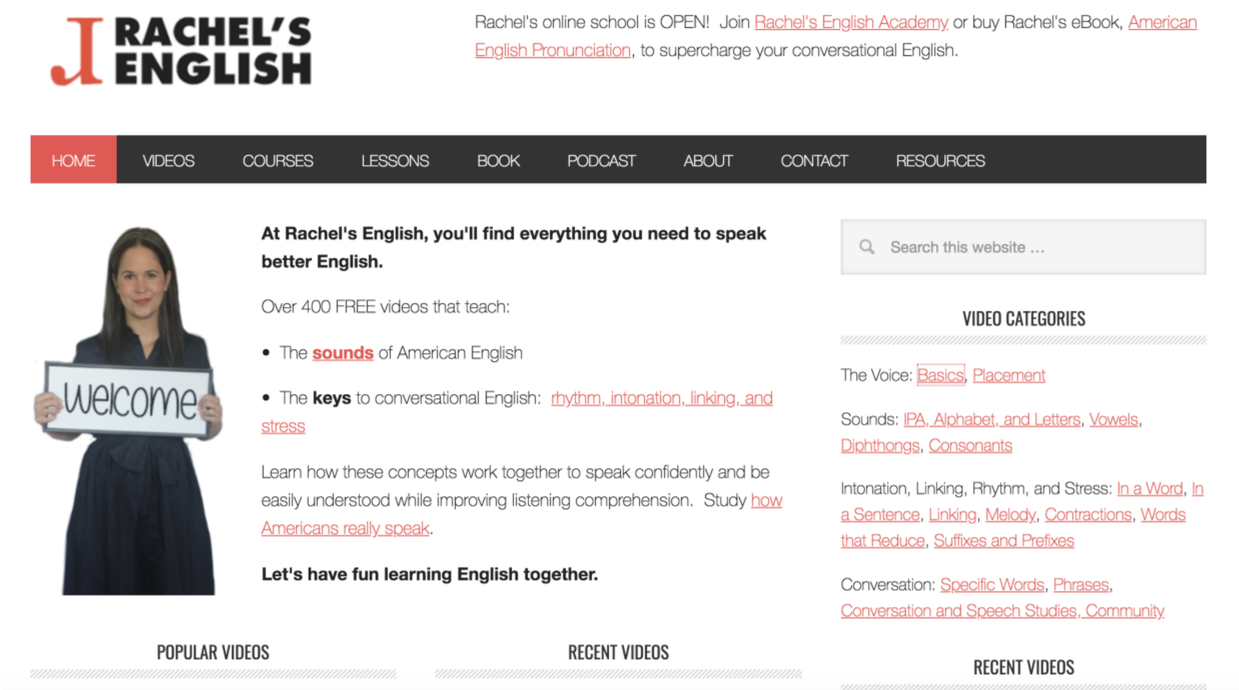


Figure 1. Rachel's English homepage.

Videos

Users are encouraged to watch the instructional videos by going to Rachel's English YouTube channel which provides themed playlists with labels such as *IPA*, *Contractions*, *Words that Reduce*, *Intonation*, and *The Dark L*. Users can also watch the videos from within the website, which is particularly helpful for learners and instructors who are located in countries where YouTube is blocked. The website presents the videos grouped thematically by category and subcategory as shown in Figure 2. For example, the category *The Voice* contains the subcategories *Basics* and *Placement*. A click on *Basics* leads users to a set of videos covering basic concepts for studying pronunciation such as the difference between voiced and unvoiced consonants, the anatomy of the voice, and exercises to relax the vocal apparatus. A click on *Placement* gives access to a set of videos that discuss where sounds are placed in American English and where the voice resonates in the body.

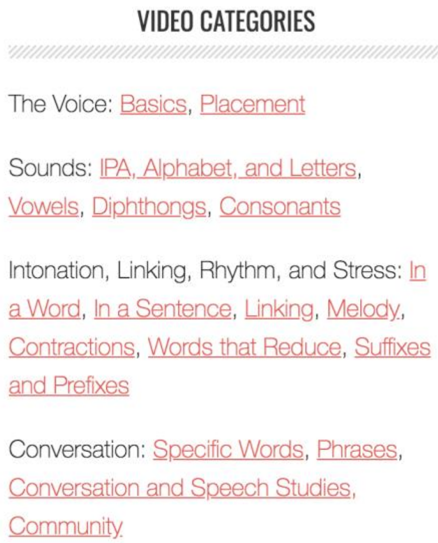


Figure 2. Video categories and subcategories.

As can be seen in Figure 3, Rachel uses video note-taking technology to demonstrate how learners can take notes on an audio or video clip. She suggests learners obtain the transcription of a spoken segment so that they can practice indicating where prominence, linking and reduction occur in a phrase or sentence. In one of the videos, Rachel uses IPA phonetic transcription and prosody annotation conventions that are typically found in pronunciation coursebooks such as Grant (2001) and Celce-Murcia et al. (2010).

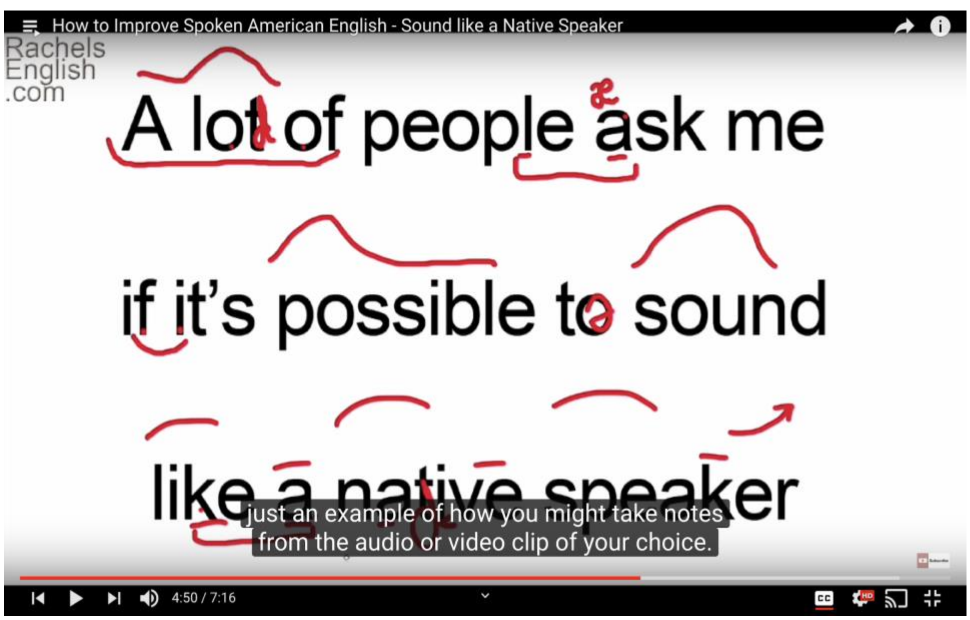


Figure 3. A demonstration of how learners can annotate prominence, linking, and reduction in transcribed text.

Another aspect of Rachel's English pronunciation teaching approach is her use of videos in which she engages in real-life conversations with friends and family members. She elicits authentic language from these interactions and she video records and transcribes them for instructional purposes. She then uses these materials to show her learners what unscripted speech sounds like and how it can be annotated and analyzed for pronunciation learning purposes. In one of the videos (see Figure 4), Rachel demonstrates how her father pronounces an individual word ("space"), an idiom ("booted out"), and a lexical bundle ("often from home") in casual conversation.

Figure 4. Pronunciation analysis of informal conversation by native speakers of American English.

One of the strategies Rachel uses to create and maintain her online learner community includes posting video challenges and encouraging learners to respond by sharing their own videos. As an example, in the Highlight of the Year challenge (see Figure 5), learners from several parts of the world shared short videos about their personal highlights.

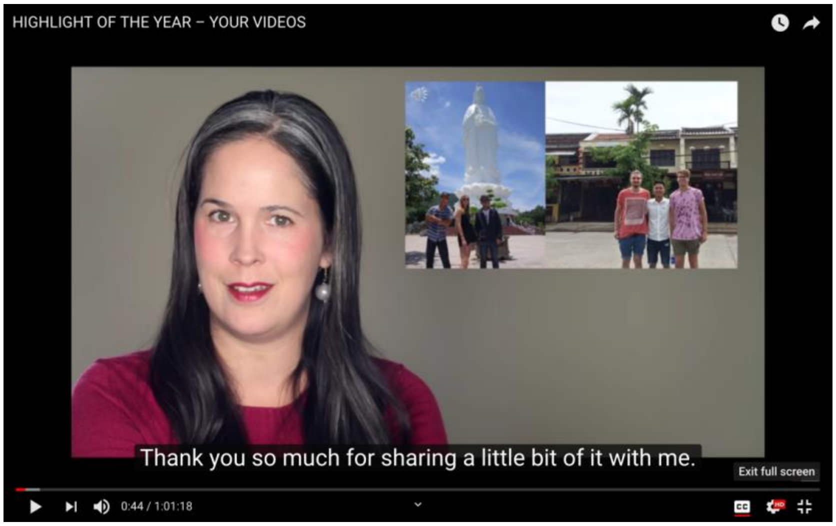


Figure 5. Learners participate by sharing videos in response to challenges posted by Rachel.

Courses

Rachel's English also offers online courses to its users. To learn more about this feature, I signed up for a course named *Accent Mini-Course 1: The Character of American English*. This is a 6-part mini-course with 3-minute videos in which Rachel explains where sounds are placed in American English and where the voice resonates in the body, which is the same content of one of the videos in the *Placement* category that was described earlier. She then asks learners to practice placing those sounds in their body by imitating her as she speaks and demonstrates. Other available courses with titles such as "International Phonetic Alphabet", "Pronunciation: Foundation", and "Vowels + Dipthongs" are available in Rachel's English Academy, which is Rachel's paid online learning website. This review, however, only describes and discusses the resources that are offered free of charge in Rachel's English.

Podcast

Rachel's English podcast has a more conversational and less instructional format. The goal of the podcast is to offer non-native English speakers the opportunity to listen to idioms, phrasal verbs and vocabulary typically used in informal conversations by native speakers of American English. The themes are varied and include conversations about American English slang, New Year's resolutions, the different pronunciations of the phoneme /t/, and interviews with English learners and instructors. Free downloadable transcriptions of each podcast are available to users.

Resources

Rachel uses her English resources page to encourage learners to seek examples of American English on the web for listening comprehension and imitation practice, and for annotating and analyzing language samples. To that end, Rachel recommends four online resources. TED Talks videos are recommended for their variety in topics and length, interactive transcripts, and subtitles in many languages. The U.S. Public TV Broadcasting Service (PBS) is highlighted for its closed captioning and extensive collection of TV programs for children. Praat, a free-downloadable voice analysis software, is recommended for learners who are interested in studying different aspects of the human voice. Finally, iTalki is recommended for learners who want to hire native speaker teachers to help them with grammar, test preparation, or business English.

EVALUATION

The freely available materials in Rachel's English have the potential to help learners and instructors of English as a second or foreign language (L2 English) to improve their speaking and pronunciation of American English. The website provides a wealth of instructional materials that learners can use for self-paced learning. Rachel's demonstrations of how users can listen to, annotate, and analyze spoken words, phrases, and sentences can equip learners and instructors with strategies and tools to explore content beyond what is available in her website. The organization of videos in categories, subcategories and playlists, and the availability of podcasts and courses can also be helpful to instructors and learners who want to focus on specific features of English speaking and pronunciation. Rachel's demonstrations of how to use IPA phonetic transcription and how to annotate prosody can also be helpful to users.

Rachel's approach to teaching pronunciation includes enunciating words, phrases and sentences so that the pronunciation features she is demonstrating sound as clear as possible. She paces herself as she explains the content, and she teaches discrete sounds in isolation at first and then as part of a larger unit of speech. This makes her speech a little slower than what would be usual for a native speaker, but it is still fast enough to be perceived as regular speech. This allows for learners to use imitation techniques such as shadowing or mirroring (Celce-Murcia et al., 2010; Derwing & Munro, 2015) to practice the sounds they hear.

L2 English learners at the intermediate or advanced level can probably make better use of the materials available in Rachel's English. However, learner participation in the video challenge activities showed that participants with limited English did engage with the materials at some level. This points to the possibility that low-proficiency learners could better benefit from the video lessons by working with L2 English instructors with some pronunciation training experience.

The resources available in Rachel's English do not offer any capabilities that allow users to interact and receive feedback on their pronunciation. Learners and instructors who seek this kind of interactive feedback would need to use Rachel's English in combination with online language learning platforms such as https://www.busuu.com/, which offer their users the possibility to interact with and provide feedback to other users.

All things considered, Rachel's English can be a valuable resource for pronunciation learning and instruction. One aspect that needs to be considered is that pronunciation courses for teachers are still not widely available, and many L2 English instructors report having insecurities on how to teach pronunciation to their learners (Burri et al., 2017; Couper, 2016). The instructional videos provided in Rachel's English can, in this sense, be used as a supplementary tool to other L2 English pronunciation instruction materials. Used in combination, these resources can help L2 English instructors work on pronunciation awareness, intelligibility, accent, accuracy, segmentals, suprasegmentals, learners' affective filters and other pronunciation issues.

ABOUT THE AUTHOR

Liberato Silva dos Santos, MA, is a PhD student in the Applied Linguistics and Technology program at Iowa State University. He researches language testing with a special focus on performance assessments of oral abilities (listening and speaking) in L2 English tests. His other research interests include investigating computer-assisted language learning (CALL) applications that can help promote student learning and instructors' professional development.

REFERENCES

Burri, M., Baker, A., & Chen, H. (2017). "I feel like having a nervous breakdown": Pre-service and in-service teachers' developing beliefs and knowledge about pronunciation instruction. *Journal of Second Language Pronunciation*, *3*(1), 109-135.

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

Couper, G. (2016). Teacher cognition of pronunciation teaching amongst English language teachers in Uruguay. *Journal of Second Language Pronunciation, 2*(1), 29-55.

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

Grant, L. (2001). *Well said: Pronunciation for clear communication*. Boston, MA: Heinle & Heinle.

Levis, J. M. (2007). Computer technology in teaching and researching pronunciation. *Annual Review of Applied Linguistics, 27,* 184-202.

Terrill, K. (2019). Pronounce Live [Review]. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Ames, IA, September 2018 (pp. 490-493). Ames, IA: Iowa State University.

TECHNOLOGY REVIEW

Pronounce Live

Kristin Terrill, Iowa State University

MEDIA SUMMARY

Pronounce Live is a mobile application that is designed to facilitate L2 pronunciation practice. The app supports 5 languages: English, French, Italian, German, and Spanish. To use this app, the user can input text as a practice item or access a pre-existing text practice item, which can be a single word or a phrase of up to 200 characters. The app generates an audio clip of text-to-speech (TTS) voice (the user can select a voice from a list of options) reading the practice item. The user then records themselves uttering the practice item. Finally, the app provides textual and graphic reports that include an automated speech recognition (ASR) transcript of their voice recording, a graphic that indicates the percentage accuracy of the user's utterance, and a report comparing their performance across multiple attempts. The interface is depicted in Figure 1.

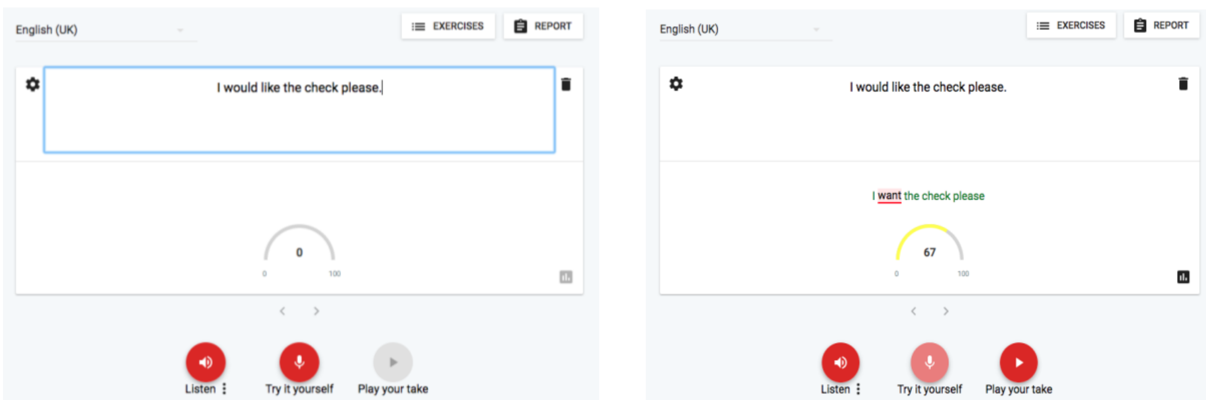


Figure 1. Pronounce Live basic practice item interface.

The target users for this application are L2 instructors and learners. Learners can use the app independently by inputting their own practice items, or by joining a group organized by an instructor. Instructors can use the app to distribute lists of practice items to their students and monitor their progress through the practice reports.

The publisher of this application is Sanako Corporation, who released the software in September, 2016. Their press release claims that with this application, "learners can practice pronunciation anywhere and anytime with the help of a tireless and always available native speech model" (Juhakoski, 2017). The software is based on their existing pronunciation product, Pronounce, which is designed to be used in a computer lab.

EVALUATION

The language learning mechanism in *Pronounce Live* is compatible with the Speech Learning model proposed by Flege (e.g. 2003), which posits that language learners' production is modeled on input. This app allows users to customize their aural input by typing the actual word or phrase that they want to learn into the interface. This feature sets *Pronounce Live* apart from other popular language practice applications on the market, which typically contain a set of pre-programmed words or phrases for users to practice, and do not permit users to add new practice items. This feature makes *Pronounce Live* a potentially useful tool for Language for Specific Purposes (LSP) teachers and learners, who could use it to encourage practicing low-frequency terminology and phraseology.

Pronounce Live also outshines other mobile language practice products in terms of the feedback it provides. Whereas popular language practice applications often provide either no feedback or undetailed feedback (for example, correct/incorrect), *Pronounce Live* provides multiple forms of feedback: an ASR-generated transcription with incorrect words emphasized, an accuracy percentage report, and the option to replay both the TTS reading of the practice item and the user's most recent recording. A feedback report that summarizes the results of each practice attempt can be downloaded as a PDF file (it is not clear whether instructors can access these reports directly through the group editor interface).

In practice, the application has some serious design flaws that impact both usefulness and usability. The app seems to function by accessing technology that is produced by other developers, such as Google. This technology includes both TTS and ASR, neither of which are designed specifically for L2 pronunciation applications. Thus, the linguistic input that the user receives is not authentic speech, but rather a computer-generated acoustic signal. Similarly, the feedback that the user receives includes a measure of accuracy in terms of percentage correct and an ASR-generated transcript. The percentage report is not very helpful, since there is no explanation provided to explain the significance of the percentage. The transcript may be useful for some, depending on their metalinguistic knowledge; however, ASR programs consider more than just pronunciation to generate transcripts (other factors include grammar and lexical frequency). Therefore, the contents of the transcript feedback may not accurately reflect real pronunciation issues. Users who attempt to trick the software or test the limits of its accuracy will find that it fails to provide feedback on some serious pronunciation problems (for example, the software often fails to differentiate between minimal pairs, especially with vowels).

Another major drawback to this application is its usability. The user interface relies heavily on graphical buttons, which is logical in an app that targets users from a range of L1 backgrounds; however, the significance of the graphical buttons is not always obvious. It is difficult to figure out how to access the text entry mode that allows users to enter their own practice items, which will certainly frustrate new users. Another usability problem relates to the list of TTS voices that users can choose from. Apparently this list is generated from metadata within the user's browser, and it includes many options with confusing labels that are listed in no particular order (see Figure 2 below).

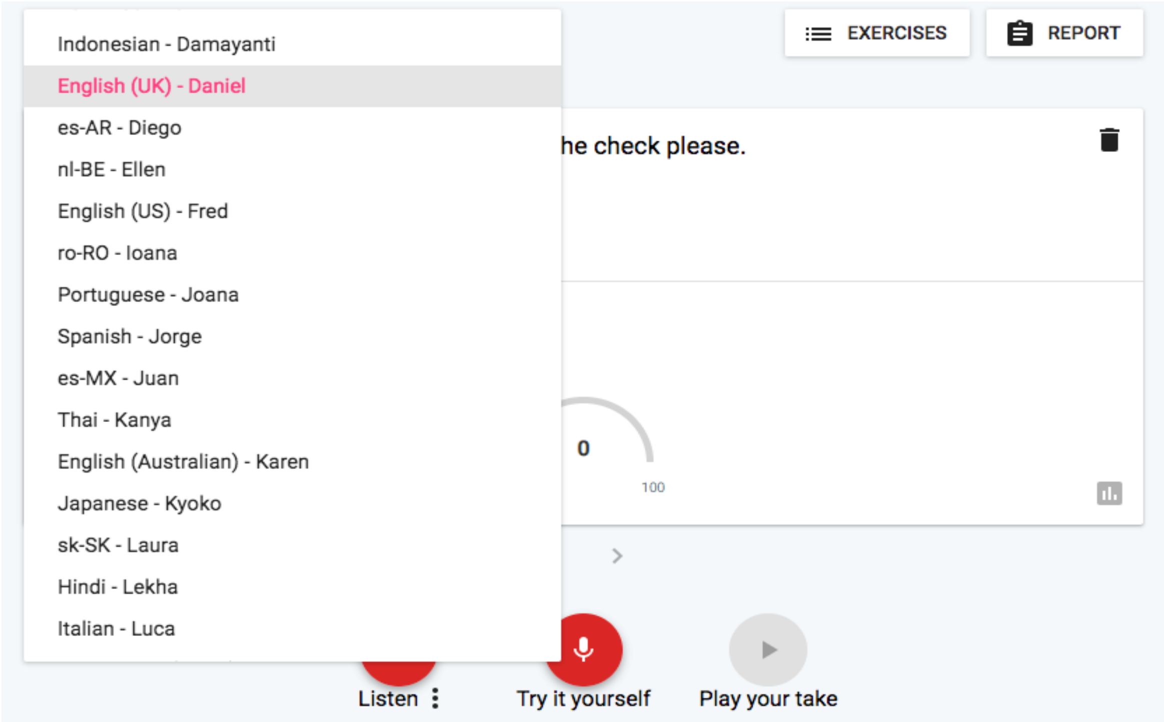


Figure 2. TTS voice selection menu.

The voices associated with the target language, in general, produce target-like examples of the practice item, but voices from other languages can be chosen to pronounce practice items. For instance, a Korean-speaking TTS voice can be selected to read an English phrase. When these options are selected, the application outputs gibberish. Because the app does not apply limits to the TTS voice option list, and no documentation explains how the list should be used, this aspect of the program is extremely confusing and negatively impacts usability. The languages supported by this application are limited and Euro-centric, which limits its usefulness as well.

RECOMMENDATIONS

Pronounce Live is an innovative concept in the field of mobile-assisted language learning (MALL). Its customizability, versatility (especially its ability to process words and phrases in 5 languages), and utility outside the traditional pronunciation laboratory make it a potentially powerful tool for language teachers and learners. Unfortunately, the problems introduced, mostly by constraints associated with TTS and ASR technology, are serious enough that many teachers will likely avoid using it. Teachers and learners who do adopt this technology should take into account the following caveats: students may have trouble using the app without ample training and in-class practice opportunities due to serious usability issues; feedback, while more detailed than most pronunciation applications, suffers from issues of accuracy and clarity; and the linguistic input provided by the application, though intelligible, is inauthentic since it is generated by a TTS program. Nevertheless, teachers might find this technology useful as a way of encouraging home practice among students and as a means of monitoring the effectiveness of home practice (using the percentage reports as a relative, not absolute, measure of students' pronunciation accuracy). One promising affordance of this application could be as a tool for practicing specialized vocabulary in an LSP context, an affordance that is not available through other MALL applications. Ultimately, *Pronounce Live* is a conceptually interesting and potentially powerful pronunciation resource that is worth continuing to develop, especially in the event that the foundational technologies, TTS and ASR, become increasingly accurate and authentic.

REFERENCES

Flege, J. M. (2003). Assessing constraints on second-language segmental production and perception. In A. Meyer, & N. Schiller (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 319-355). Berlin: De Gruyter Mouton.

Juhakoski, A. (10 April 2017). Sanako Corporation Press Release. Retrieved from http://www.sanako.com/perfect-your-pronunciation/