PROCEEDINGS OF THE 8TH ANNUAL PRONUNCIATION IN SECOND LANGUAGE LEARNING AND TEACHING CONFERENCE

(ISSN: 2380-9566)

Editors

Mary O'Brien

John Levis

PSLLT 2016 CONFERENCE PROGRAM

8th Annual Pronunciation in Second Language Learning and Teaching Conference

The Role of Technology in L2 Pronunciation Research and Teaching

University of Calgary

August 12–13, 2016

PSLLT 2016 Conference Schedule

Thursday, August 11

5:30-8:00 pm	**Blue Room, Dining Centre** Welcoming Reception *The Dining Centre is west of Hotel Alma. See the map at the end of the conference program.*

The conference will take place in the Rozsa Centre on the University of Calgary campus.

Friday, August 12

8:00-10:00	**Lobby, Rozsa Centre** On-site registration and coffee		
8:45-9:00	**Great Hall** Conference Welcome		
9:00-10:00	**Great Hall** *Plenary* ***John Levis (Iowa State University, USA):*** "Technology and the intelligibility-based classroom"		
10:00-10:30	Coffee break		
	Great Hall *Assessment* Chair: J. Crabtree	**Hub Room** *Teaching suprasegmentals* Chair: M. Richards	**Milling Room** *L2 Accent* Chair: T. Barriuso
10:30-10:55	**Munro:** An empirical comparison of comprehensibility scaling techniques	**Dickerson:** The baby in the rhythmic bathwater	**Koffi:** An acoustic phonetic account of the confusion between [r] and [l] in seven varieties of L2 English
11:00-11:25	**Piccardo & North:** Assessment scales for phonological control: creating and validating new descriptors for the CEFR	**Saito:** Reexamining the differential effects of instruction on English suprasegmentals	**Pérez Ramón, García Lecumberri & Cooke:** Linguistic competence and the perception of degrees of foreign accent

11:30-11:55	**Menjo:** Using PEPS-C to investigate L2 prosodic acquisition of adult Japanese learners	**He & Wang:** Approaches to Mandarin tone 3 teaching in classroom: Contour tone vs. low tone	**Tronnier & Kupisch:** L1 transfer in the acquisition of gemination in L2 Italian
12:00-12:25	**Foote:** I can't understand you because of…um…your pronunciation: Analysing learner descriptions of pronunciation in comprehensibility judgments		**Simonchyk & Darcy:** Lexical encoding and perception of palatalized consonants in L2 Russian
12:30-2:00	**Evans Room** Poster session and lunch (provided)		
	Great Hall *Orthography* Chair: T. Derwing	**Hub Room** *Computer-assisted pronunciation training* Chair: I. Martin	**Milling Room** *L2 pronunciation teaching* Chair: K. Yokomoto
2:00-2:25	**Gomez Lacabex & Garcia Lecumberri:** The effect of orthography in the production of English schwa by Spanish speakers	**Bliss, Cheng, Lam, Pai, Schellenberg & Gick:** Ultrasound technology and its role in Cantonese pronunciation teaching and learning	**Shehata:** Teaching Arabic pronunciation to non-natives: cognition, practice and outcomes
2:30-2:55	**Barrios & Hayes-Harb:** L2 learning of phonological alternations with and without orthographic input	**White, Gananathan & Mok:** Teaching dark /l/ with ultrasound technology	**Sturm:** Phonetics instruction and the /u/-/y/ distinction in French as a foreign language: A preliminary study

3:00-3:25	**Showalter:** Inhibition with incongruent grapheme-phoneme correspondences in English speakers learning Russian	**Qian, Chukharev-Hudilainen, & Levis:** A self-access computer-assisted HVPT application for learner diagnostic and perceptual training on English segmentals	**Miyamoto, Suzuki, Sturm, & Fukada:** Survey of the views on foreign language pronunciation: Comparing students' and teachers' beliefs
3:25-3:45	Break		

	Great Hall *Listener factors* Chair: A. Shehata	**Hub Room** *Computer-assisted pronunciation training* Chair: J. Trimble	**Milling Room** *Linguistic factors* Chair: S. Menjo	**Evans Room** *Pronunciation development* Chair: A. George
3:45-4:10	**Hayes-Harb & Showalter:** Linguistically-guided adaptation to accented speech: The roles of lexical frequency and word familiarity	**Wang:** L2 Pronunciation Training: Comparisons of English vowel and Mandarin tone training methods and results	**Zhou & Li:** Can fluency predict raters' decisions? Exploring the relationship between fluency measures and speaking performance of prospective International Teaching Assistants	**Liu:** Mandarin speaker's emphasis in English and Mandarin lectures
4:15-4:40	**Kang & Johnson:** Measures of intelligibility in different varieties of English: Human vs. machine	**Yang:** Perception assessment versus contour assessment for online tone training	**Thompson & Huensch:** Pronunciation attitudes: The role of multilingual status and perceived positive language interaction	**Kanefsky & Violin-Wigent:** The role of explicit instruction in the acquisition of liaison in French: The intersection of syntax and phonology

4:45-5:10	**Ghanem:** Nonnative speakers' use of linguistic features with different interlocutors	**Isbell, Lee & Park:** Technology-integrated pronunciation instruction for learners of Korean	**Zetterholm & Falk:** L1 as a transfer source for pronunciation in L3 learning	**Hernandez-Morales:** Development in pronunciation accuracy through visual feedback and repetition: Evidence from SRC
6:00-9:00	**Westin Hotel -Downtown** (320 4th Avenue SW, Britannia Room) Conference Dinner *Note: At 5:30 graduate students will escort participants from the conference venue to the C-Train and then to the hotel. Fare is $3.15 each way.* Directions: *Travel south on the train to the 3rd Street SW stop. From there it is a 6-minute walk north along 3rd Street for three blocks. The hotel is on the right.* *Taking a taxi from the conference venue is another option. The trip would probably cost about $15 each way.*			

Saturday, August 13

8:30-9:00	Coffee		
9:00-12:00	**Great Hall** *Roundtable Discussion: Role of technology in L2 pronunciation teaching and research* **Catia Cucchiarini:** The digital literacy instructor: ASR for adult L2 literacy students **Deborah Hardison:** Visualizing the gesture and prosodic components of emphasis in oral discourse **Hansjörg Mixdorff:** Prosodic aspects of foreign accent with a view to CAPT Break **Ron Thomson:** High variability phonetic training: Recent insights, current limitations, and future potential Respondent: Mary Grantham O'Brien Question and answer period Chair: Jennifer Foote		
12:00-1:30	Lunch (on own) *JSLP* editorial board will meet in the Milling Room.		
	Great Hall *Teacher education* Chair: S. Sonsaat	**Hub Room** *Computer-assisted pronunciation training* Chair: S. Rossi	**Milling Room** *Phonetics* Chair: H. Bliss
1:30-1:55	**McGregor, Meyers, Reed, Zielinski & Liu:** Navigating the terminological minefield of intonation: from theory to research to textbooks	**Butler:** Understanding dialect diversity through spectrographic comparison	**Sakai:** Channeling Frankenstein: How to synthesize vowels in *Praat*
2:00-2:25	**Yokomoto:** An instructional decision making model for EFL pronunciation teaching	**George:** Teaching Spanish intonation variation using technology	**López-Morelos, Llama & Méndez:** VOT production by English-speaking Spanish heritage speakers enrolled in French immersion

2:30-2:55	**Buss:** The role of training in shaping pre-service teacher cognitions related to pronunciation	**Trimble:** Technology-enhanced teaching of Spanish vowels and its effect on vowel pronunciation in introductory Spanish courses	**Barriuso & Barrios:** The role of distributional information in the acquisition of a novel allophonic alternation
3:00-3:30	Break		
3:30-5:00	**Great Hall** Tips and Strategies for Teaching Pronunciation – Round Robin format; Chair: Wendy Chambers, ATESL		

TEACHING TIPS AND STRATEGIES

PRONUNCIATION IN SECOND LANGAUGE LEARNING AND TEACHING
8th Annual Proceedings
(Conference held August 2016, Calgary, Alberta, Canada)

TABLE OF CONTENTS

TEACHING TIPS

REVIEWS

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*, ISSN 2380-9566, Calgary, AB, August 2016 (pp. 1-9). Ames, IA: Iowa State University.

PRONUNCIATION AND TECHNOLOGY

Mary G. O'Brien, University of Calgary
John M. Levis, Iowa State University

INTRODUCTION

Pronunciation training that makes use of technology has come a long way from its early days in language laboratories equipped with record players. The current explosion in new technologies means that language learners are now capable of working on their pronunciation at any time, regardless of where they are. Web-based programs and mobile apps that claim to improve learners' pronunciation are readily accessible, and most are relatively inexpensive. Nonetheless, many of the commercially available products are often neither pedagogically sound nor informed by research (Foote & Smith, 2013). There is clearly a need for collaboration among pronunciation researchers, software developers, and classroom language teachers to determine which aspects of pronunciation should be prioritized for which types of learners, which types of pronunciation activities are most beneficial for developing pronunciation skills, and how these technologies can best be used to enhance classroom teaching. Most importantly, the goal of any new pronunciation technology should be the development of more intelligible speech. That is, pronunciation training should enable second language (L2) learners to be more easily understood (Levis, 2005).

The use of pronunciation technologies within an L2 teaching and learning context is captured within the field of Computer-Assisted Pronunciation Teaching (CAPT). Recent reviews of the CAPT literature (Chun, 2013; Levis, 2007; O'Brien, 2011) often demonstrate promising results, especially with software that provides learners with specific feedback on their errors. Researchers tend to rely on software that was developed for the recording and analysis of learners' speech. On the other hand, language learners and teachers are often on the lookout for user-friendly software that has been developed specifically for the purposes of student training (i.e., courseware). Until recently, the gulf between the use of technology for research and teaching has seemed far too wide. Although software used in research laboratories like *Praat* (Boersma & Weenink, 2017) and *Audacity* (Audacity Team, 2017) have been developed with the goals of researchers in mind, with training and a set of clear learner tasks, this same technology can be used for teaching purposes (e.g., Hardison, 2004; Levis & Pickering, 2004). Similarly, courseware created for the purposes of training learners' perception (e.g., *English Accent Coach,* Thomson, 2017) or production (e.g., *DISCO*, Strik, Colpaert, van Doremalen, & Cucchiarini, 2012) can be used to gather data on pronunciation development. Moving forward, collaborations among stakeholders will result in more research-informed pronunciation technologies that have the capability to truly improve learners' pronunciation in ways that are tailored to individual learners' needs.

Encouraging collaboration among stakeholders was a primary goal of the 2016 Pronunciation in Second Language Learning and Teaching (PSLLT) Conference. The theme of the Conference, held at the University of Calgary on August 12-13, 2016, was

"The Role of Technology in L2 Pronunciation Research and Teaching." It featured 41 individual papers and 13 posters, a plenary by John Levis on the use of technology in the intelligibility-based classroom, 11 research-based tips and strategies for teaching pronunciation, and a roundtable on the development and use of cutting-edge technology in L2 pronunciation teaching and research. Sponsored by *Language Learning*, the roundtable featured the work of Debra Hardison, Catia Cucchiarini, Hansjörg Mixdorff, and Ron Thomson. The organizers were fortunate to receive substantial funding from other sponsors including the Social Sciences and Humanities Research Council of Canada (SSHRC), the University of Calgary Faculty of Arts, the University of Alberta Faculties of Education and Extension, and the Alberta Teachers of English as a Second Language (ATESL).

We are pleased to share the Proceedings with you. A representative sample of conference papers and teaching tips and strategies are included here. Many of the contributions feature specific technologies including sophisticated hardware (i.e., ultrasound), acoustic analysis software like *Praat*, and software designed for purposes other than teaching pronunciation (e.g., *Qualtrics*, *Anytune*). The themes of the papers extend beyond the use of technology and include the assessment of L2 pronunciation, the effects of pronunciation training, the ability of learners to perceive and / or produce segmental differences in their L2, prosodic features of the L2, and learner and teacher beliefs and practices. We have also included nine website/software/corpus/book reviews. Although these reviews were not presented at the conference, they align well with the conference theme and will be of special interest to pronunciation researchers and teachers. Below we present a brief summary of each of the contributions.

PAPERS

In their paper "The Role of Phonological Distributional Information on the Acquisition of L2 Allophones", Taylor Anne Barriuso and Shannon Barrios focus on adults' ability to distinguish between [b] and [β] after exposure to these segments in an artificial language. The authors were primarily interested in determining whether participants exposed to the sounds in overlapping contexts (i.e., the segments were presented in all of the same contexts) would differ in their sensitivity to these sounds from participants who heard the sounds in complementary contexts (i.e., the contexts did not overlap). Although they expected that listeners would be able to make use of distributional information and that participants in the overlapping context would be better able to distinguish between [b] and [β] on an ABX task, they found no such pattern, and they concluded that the participants were unable to infer the phonological status of the two segments.

Shannon Becker examines the ability of English-French L2 learners from three separate levels of French instruction to perceive and produce the French nasal vowels /ɑ̃/ and /ɔ̃/ in her contribution "Perception and Production of Unfamiliar L2 Segments: Using Technology for Teaching and Research." The results demonstrate that learners did not differ according to class level and that there were no changes from pre- to posttest. Participants showed a great deal of variation in their production of these segments. The final section of the paper encourages instructors to make use of research tools like *Praat* and *Qualtrics* survey software to create perception and production exercises for students.

The contribution "Ultrasound Technology and its Role in Cantonese Pronunciation Teaching and Learning" by Heather Bliss, Lauretta Cheng, Murray Schellenberg, Zoe Lam, Raymond Pai, and Bryan Gick focuses on the use of ultrasound overlay videos as a pronunciation teaching tool in beginner level university Cantonese language classes. The researchers taught the pronunciation of two sets of sounds (unreleased final stops and low central vowels) to learners under two conditions. One group of participants used the overlay videos and the other used audio-only materials. The results point to better performance in perception and production among learners who made use of the videos. Based on the findings, the authors propose that such videos could be used as pronunciation teaching tools in a range of languages.

Wayne B. Dickerson's piece, "The Baby in the Rhythmic Bathwater," proposes an alternative to teaching English stress-timed rhythm, which he calls the two-peak profile. Although he notes that there are certain aspects of pedagogical materials that rely on a model of stress-timed rhythm that are worth using in the classroom, he posits that the two-peak profile is both better aligned with the rhythm that is employed in spontaneous speech. His contribution demonstrates how teachers can integrate the two-peak profile approach even when their classroom resources are based on a model of stress-timed rhythm.

David O. Johnson and Okim Kang, in "Measures of Intelligibility in Different Varieties of English: Human vs. Machine," introduce an automated tool to measure the intelligibility of English speech. The performance of the tool was compared to human measure of intelligibility of six varieties of English speech (American, British, Indian, South African, Chinese, and Spanish). The computer tool was used to identify up to 11 features that likely affect intelligibility scores.

In his paper "An Acoustic Phonetic Account of the Confusion between [ɹ] and [l] in Seven Varieties of L2 English: Focus on Intelligibility and Accentedness," Ettien Koffi provides analyses of [ɹ] and [l] productions of English L2 learners from a range of first languages (L1s) and compares them to those produced by native speakers of American English. The results of acoustic analyses of F3 values and vibration rates of the consonants demonstrate that some participants produced /l/s and /ɹ/s that are indistinguishable from one another. Koffi concludes with pedagogical implications that focus on encouraging intelligible production of the consonants.

Di Liu's "A Mandarin Speaker's Intonational Emphasis in English and Mandarin Lectures" examines one Mandarin-English L2 learner's prosodic marking of new and old information in an identical lecture given in both of his languages. Specifically, Liu looks at the speaker's use of maximum pitch to highlight new constituents. The results indicate that the speaker uses pitch to contrast new and old information to a greater extent in Mandarin than in English. Liu proposes that learners should be encouraged transfer prosodic features from their L1 to their L2, even when the two languages are typologically dissimilar.

Enrica Piccardo and Brian North report on their project aimed at developing new descriptors targeting pronunciation for the Common European Framework of Reference in their paper "Developing Phonology Descriptors for the Common European Framework of Reference (CEFR)." The authors describe the multi-staged consultation process and the qualitative and quantitative analysis of data that they carried out to come up with the new descriptors of phonological proficiency and new scales (sound articulation and prosody). Whereas the previous scale of phonological control was based on notions of accentedness, the authors point to the central role played by intelligibility in the new descriptors.

Asmaa Shehata's "Teaching Arabic Pronunciation to Non-Natives: Cognition and Practice" examines the extent to which two Arabic teachers' beliefs about pronunciation teaching are borne out in their classroom practice. Shehata probed alignment between teachers' and students' beliefs via questionnaires and interviews and actual practice via classroom observations. In general, the teachers reported difficulty with teaching pronunciation, and they relied primarily on controlled activities that focus on segmentals when they did teach pronunciation. The results of the study point to a need for language-specific pronunciation training for language teachers.

In their paper "Lexical Encoding and Perception of Palatalized Consonants in L2 Russian," Ala Simonchyk and Isabelle Darcy report on their study investigating English-Russian L2 learners' perception of Russian of the plain/palatalized contrast in the /l/ vs. /l^j/ pair. Their goal was to determine whether there is a relationship between participants' abilities to perceive the contrast in an ABX task and their ability to encode and retrieve words with the contrast in an auditory word-picture matching task. Although they found no relationship between the scores on the perceptual and encoding/retrieval tasks among intermediate learners, Simonchyk and Darcy demonstrated a strong relationship between the performance on the two tasks for advanced learners of Russian. This leads the authors to conclude that there is a strong connection between learners' perception and lexical encoding of the contrast.

Jessica Sturm investigates the effectiveness of training on English-French L2 learners to produce the /u/-/y/ contrast in "Phonetics Instruction and the /u/-/y/ Distinction in French as a Foreign Language: A Preliminary Study." Participants came from two groups: one group received explicit instruction on the contrast in the context of a French phonetics and pronunciation course, and the other group did not. Although Sturm did not find significant acoustic differences in the production of the vowels between the trained and untrained groups of learners after training, she suggests that earlier and/or more systematic training might have a more profound impact on learners' ability to distinguish between the vowels.

Amy Thompson and Amanda Huensch explore the relationship between learners' status as bilinguals/multilinguals and their attitudes toward improving their pronunciation in their contribution "Pronunciation Attitudes: The Role of Multilingual Status and Perceived Positive Language Interaction (PPLI)." They operationalized multilingualism in two ways: traditionally (i.e., learners' self-reports of number of languages spoken) and as PPLI (i.e., multilingual learners are those who perceive positive interactions between their languages). The results of the study demonstrate that multilingual and PPLI participants demonstrated a stronger desire to improve their pronunciation than bilingual participants did. Thompson and Huensch argue that language learners' multilingualism should be viewed as an asset in the language classroom and that instructors should inform learners about the benefits of crosslinguistic interactions.

Donald White, Richard Gananathan, and Peggy Mok report on the results of a training study in their contribution "Teaching Dark /l/ with Ultrasound Technology." The participants in their study, eight Cantonese-English L2 learners, read sentences containing [ɫ] before and after training. One group received feedback from the ultrasound scanner on their productions, and the other did not. The authors found that most of the participants who received feedback improved, but they did not find evidence for improvement among the participants in the no-feedback condition. They therefore conclude that even very short ultrasound training may be an effective way to teach [ɫ] to Cantonese-English L2 learners.

In their paper "Exploring the Relationship between Fluency Measures and Speaking Performance of Prospective International Teaching Assistants," Ziwei Zhou and Zhi Li investigate the extent to which four categories of fluency measures (i.e., speed, juncture pauses as breakdown, non-juncture pauses as breakdown, and fillers) predict the oral proficiency scores assigned to the speech samples of International Teaching Assistants (ITAs). The results of a multiple regression analysis point to average syllable duration and juncture pauses as the best predictors of overall proficiency scores. The findings add to the growing body of literature investigating the ability of automated systems to evaluate L2 learners' speaking proficiency and may inform L2 speaking proficiency assessment, both in terms of rating schemes and assessor training.

TEACHING TIPS

For the third year, we include Teaching Tips in the proceedings. Teaching Tips are done at the conference in a Round-Robin format. For 8-10 minutes, presenters teach their tip to a table of participants. At the end, participants go to another table, and presenters have a new set of participants to present to again. In a 90-minute period, participants have the opportunity to try about 8 teaching tips. The weakness of this system is that presenters don't get to see other presenters, but the energy level of the session is amazing.

In his teaching tip "Oye mi Canto, mi Son: Using Tongue Twisters and Songs," Douglas Bowman presents a series of activities that he has used to teach Spanish <o> to beginner level middle school and high school learners of Spanish. He focuses specifically on the use and repetition of tongue twisters and songs in order to encourage the solidification of sound-symbol correspondences.

Marsha Chan's teaching tip, "*Anytune* Slows Down Sound Tracks for Language Practice," provides clear instructions for how language learners and instructors may use the slow-downer app *Anytune* to slow down sound files without changing the pitch of the original sound file. She highlights the ability of the software to speed up the sound file incrementally so that learners can control the tempo and can listen to a speech sample at a particular tempo as often as they wish.

Brenda Imber, Carson Maynard, and Maria Parker make the case for using visualization cues in *Praat* to improve L2 learners' comprehensibility in their teaching tip, "Using *Praat* to Increase Intelligibility through Visual Feedback." The starting point for their contribution is their experience teaching graduate student ESL oral skills courses. They note that learners' ability to evaluate their pronunciation and see their improvement enables them to develop autonomy through in-class work and homework assignments. In the tip, the authors lay out their progression for training students how to use *Praat*, and they provide evidence of one student's suprasegmental improvement after seven weeks of instruction.

In her teaching tip, "Teaching Pronunciation through Homework Assignments: The Method of iCPRs", Ines Martin demonstrates how teachers can make use of innovative Cued Pronunciation Readings (iCPRs) to teach pronunciation. The iCPRs, which can be created by instructors using *PowerPoint*'s built-in features, follow the same progression: perceptual training (accentedness detection followed by sound discrimination) and production training. Martin demonstrates that iCPRs assigned as homework in both face-to-face and online education settings effectively target pronunciation.

Elizabeth Zetterholm's contribution "Teaching the Pronunciation of Swedish Exotic Vowels" outlines steps that teachers can take when teaching new L2 vowels. Her teaching tip focuses primarily on teaching Swedish rounded vowels, which are relatively rare in the languages of the world and which have been shown to cause difficulties for L2 learners. She notes that instructors should focus primarily on the articulatory differences among the vowels. In order to do so, she proposes that language learners can use mirrors to enable observations of their jaw openings and pencils above their upper lips to encourage lip protrusion.

REVIEWS

Similar to last year, we add a set of reviews of websites, books, and software. These were not part of the conference, but were developed by John Levis's graduate students in a course on Technology and Oral Language at Iowa State University during Fall semester 2016. We share them in the proceedings both because of their interest to readers of the proceedings and because of their relevance to the conference theme.

Mo Chen reviews *Saundz*, an app designed for English-as-a-foreign-language (EFL) students which offers computer assisted pronunciation training. The goal of the software is to help students learn American English sounds, give them easy access to pronunciation tutoring, and help reduce their accents. The review evaluates the main features of Saundz according to a much-used CALL evaluation framework. While the app is useful on a word-by-word basis, it could be improved with the addition of visuals, a more individualized approach to feedback, and a more meaningful context.

Idée Edalatishams reviews the *LeaP corpus*, a collection of speech from L2 learners of German and English. The context of the corpus creation was investigating the acquisition of prosody at phonetic and phonological levels. The review reveals that while some prosodic features are included in annotations, more should be added for pitch, and that the annotations also are unreliable at times. Additionally, the corpus is less user friendly than is desirable, but this may be a function of the age of the corpus design.

Nazlinur Gokturk looks at how *Mondly* uses gamification to present vocabulary and conversational skills in 33 languages. The app presents 23 different situations a user is likely to run into, and presents 6 relevant lessons on vocabulary and conversational skills pertaining to that situation. While these lessons are good for a novice learner, the conversation lessons are limited enough to not be as beneficial for a higher level learner. An improvement that could be made would be to include an explicit feedback system.

Voicetube allows students to practice pronunciation through videos, and it is reviewed by Haeyun Jin. *Voicetube* is a Taiwan-based web application and is accessible through a main website, iOS and Android mobile applications. An extra feature *Voicetube* provides is a shadowing tool in the speaking section of the videos. This gives students the ability to practice shadowing through listening, speaking, and vocabulary. While the content is authentic and the site provides highly individualized practice, there is room for improvement. Some of these include adding an automatic speech recognition feature to provide feedback, more levels of the shadowing feature, and other technical limitations.

Yasin Karatay reviews the multiple functions provided by *YouGlish,* a searchable video database, which provides short video segments for any word or phrase contained within the database. When users sign up for an account, the site provides a lesson of the day, a word of the day, and users are able to save videos to a set location. While there are many benefits to this easily searchable database of speech, there are also a few drawbacks to the site. *YouGlish* is output based, and there is no way to record speech to receive feedback on pronunciation. Also, a phonetic transcription of the words is not provided.

Jeremy Lockwood reviews *NORM: The Vowel Normalization and Plotting Suite*, a website developed for socio-phoneticians, phoneticians, and sociolinguists to facilitate manipulation, normalization, and plotting of vowel formants. The site assumes its users have sufficient background knowledge in acoustic analysis and acquiring formant data. The site is difficult to navigate, and researchers without the proper background knowledge cannot easily use it. For those with the ability, however, NORM provides excellent ways to visualize acoustic vowel data.

The *American English Pronunciation Tutor* smartphone application, which offers ten units covering segmentals and suprasegmentals in English, is reviewed by Sock Wun Phng. One major benefit of this app is its focus on developing production and perception without pushing the goal of accent reduction. The app includes an orthographic representation of the target sounds and IPA symbols. The activities are repetitive, and the app could be improved with the incorporation of game-based learning theory. Giving the activities more focus on meaning would also increase the appeal of the app.

Alif Silpachai reviews *Accent Reduction for Professionals: How to Eliminate Your Accent to Sound More American*, which focuses on accent reduction. Overall, the book is disappointing and shows little understanding of generally accepted research findings. It is also misleading, and encourages speakers to lose their accents in order to avoid discrimination in the workplace, leading to often confusing and contradictory recommendations.

Finally, Taichi Yamashita discusses *Manythings.org,* a Japan-based website designed for learners who wish to study American English independently. This website has several benefits including providing feedback for the learner and repetition in the minimal pair section of the lessons. However, learners do not have the opportunity to produce language, but only to read and listen. The website also focuses on input with less importance placed on meaning or form.

CONCLUSION

The PSLLT Proceedings are now in their 8th year, and the conference in its 9th year. The proceedings have provided a venue for around 200 articles, teaching tips and reviews that are freely available. Many other presentations at the conference have been published in refereed journals. Since 2009, the number of professional books on L2 pronunciation has exploded, there is a dedicated *Journal of Second Language Pronunciation*, several other conferences on L2 pronunciation are regularly scheduled (Accents, English Pronunciation: Issues and Practices, The Phonetics Teaching and Learning Conference, New Sounds, etc.), and L2 pronunciation as a field has expanded well beyond English to include L2 pronunciation of a wide variety of languages. This is exciting, and we look forward to seeing what the next years bring as the field continues to take shape.

REFERENCES

Audacity Team. (2017). Audacity. Version 2.1.0, retrieved 18 March 2017 from https://sourceforge.net/projects/audacity/

Boersma, P., & Weenink, D.(2017). *Praat*: Doing phonetics by computer. Version 6.0.26, retrieved 18 March 2017 from http://www.fon.hum.uva.nl/praat/

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

Foote, J., & Smith, G. (2013, September). *Is there an app for that?* Paper presented at the 5th Pronunciation in Second Language Learning and Teaching Conference, Ames, Iowa.

Hardison, D. M. (2004). Generalization of computer-assisted prosody training: Quantitative and qualitative findings. *Language Learning & Technology 8*(1), 34–52.

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

Levis, J. (2007). Computer technology in teaching and researching pronunciation. *Annual Review of Applied Linguistics, 27,* 184-202.

Levis, J. & L. Pickering (2004). Teaching intonation in discourse using speech visualization technology. *System, 32*(4), 505–524

Motohashi-Saigo, M. & D.M. Hardison (2009). Acquisition of L2 Japanese geminates: Training with waveform displays. *Language Learning & Technology, 13*(2,) 29–47. http://llt.msu.edu/vol13num2/motohashisaigohardison.pdf.

Strik, H. Colpaert, J., van Doremalen, J., & Cucchiarini C. (2012). The DISCO ASR-based CALL system: practicing L2 oral skills and beyond. *Proceedings of the Conference on International Language Resources and Evaluation (LREC 2012),* Istanbul, May 2012.

Thomson, R. I. (2017). *English Accent Coach* [Computer program]. Version 2.3. Retrieved from www.englishaccentcoach.com

Barriuso, T.A. and S. Barrios. (2017). The role of phonological distributional information on the acquisition of L2 allophones. 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. 10-20). Ames, IA: Iowa State University.

THE ROLE OF PHONOLOGICAL DISTRIBUTIONAL INFORMATION ON THE ACQUISITION OF L2 ALLOPHONES

Taylor Anne Barriuso, University of Utah
Shannon Barrios, University of Utah

It has been well attested that infants and adults are able to take advantage of statistical distributional information to acquire phonemes (Maye et al., 2002; Hayes-Harb, 2007) and that infants can learn novel phonological alternations on the basis of phonological distributional information (White et al., 2008). Less is known about the way in which adult second language (L2) learners acquire allophonic relationships. The present study investigates the role of a phonological distributional mechanism in a controlled experimental context. We asked whether naïve subjects were able to utilize phonological distributional information to determine whether two phones belong to separate phonemes or a single phoneme. We exposed native English speakers to one of two artificial languages in which two acoustically similar sounds ([b] and [β]) occurred in either overlapping or complementary distribution. After the exposure phase, participants completed an ABX discrimination task. Unexpectedly, participants did not perform differentially on the task depending on their exposure type, failing to provide evidence for the use of a phonological distributional mechanism in adult L2 allophonic acquisition.

INTRODUCTION

When acquiring the phonology of a language, learners must learn not only the contrastive sound categories, but also the positional variants (allophones) associated with those phonemes. Child language acquisition researchers have determined the timeline over which native language (L1) categories are acquired: vowels by 6 months (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 2006), most consonants by 8-10 months (Werker & Tees, 1984), and allophones by about 10-12 months of age (White, Peperkamp, Kirk, & Morgan, 2008). Several learning mechanisms have been identified to account for this rapid native phonological acquisition. Infant learners have been shown to be able to leverage a statistical distributional learning mechanism (Maye, Werker, & Gerken, 2002) as well as various lexical mechanisms, such as distinct visual referents (Yeung & Werker, 2009), and distinct word forms (Feldman, Myers, White, Griffiths, & Morgan, 2013) to support phonetic category learning. A phonological distributional mechanism has also been proposed for the learning of phonological status (Peperkamp, Pettinato, & Dupoux, 2003; Peperkamp, Le Calvez, Nadal, & Dupoux, 2006).

Research on adult L2 learners has also extensively investigated the acquisition of novel phonemic contrasts finding that numerous factors have an influence, including the relationship between L1and L2 phonetic and phonological systems, experience factors including age of acquisition and length of residence, and typological factors, among others (Flege, 1995; Best, 1995; Best & Tyler, 2007; Eckman, 1981). Evidence suggests that statistical distributional and lexical mechanisms may still be available for L2 phonological acquisition (Hayes-Harb, 2007). The present study investigates whether phonological distributional mechanisms, those thought to account for infant acquisition of allophone categories (Peperkamp et al., 2003; Peperkamp et al., 2006; White et al., 2008), may also be available for the acquisition of allophone categories by adult L2 learners.

BACKGROUND

Traditionally phonological theory has made a distinction between two levels of representation—phonemes and allophones—which differ in their phonological distributions. Phonemes occur in overlapping distribution; they occur in the same phonological environments as other phonemes and are, therefore, not predictable by their context. Allophones occur in complementary distribution; different allophones of the same phoneme do not occur in the same phonological environments. Instead, the phonological context conditions which allophone occurs. Categories at each of these levels of representation differ cross-linguistically (e.g. [d] and [ð] map to a single phoneme category in Spanish but to two different categories in English), so both L1 and L2 learners must use information available in their input to learn them. The properties of the input required to do this are not yet determined, although various learning mechanisms have been proposed for native phonological acquisition.

Maye et al. (2002) provided evidence for a statistical distributional learning mechanism—the ability to track the frequency of particular phones in acoustic-phonetic space in order to develop phonemic categories. Lexical mechanisms—the ability to leverage distinct word forms, visual referents, and word-meaning pairs to bootstrap the formation of distinct phonetic categories—can also be used by infant language learners (Yeung & Werker, 2009; Feldman et al., 2013). Others have provided evidence that phonological distributional mechanisms—the ability to track phones with regard to their phonological environment—contribute to word segmentation, parsing, and allophone learning (Saffran, Aslin, & Newport, 1996; Chambers, Onishi, & Fisher, 2003). Importantly, White et al. (2008) provided evidence that infants can use a phonological distributional learning mechanism to group allomorphic alternants into the same category.

Despite the differences that have been noted between L1 and L2 acquisition, evidence suggests that many of the learning mechanisms proposed for child language acquisition may also be available for and leveraged by adults. Hayes-Harb (2007) asked whether adult second language learners were able to use statistical and lexical information to learn a novel phonemic contrast. The results provided evidence that adult learners were able to use information about the acoustic-phonetic distribution of phones to learn novel contrasts. They were also able to use lexical information in the form of distinct visual referents to learn a novel contrast, suggesting that both statistical and lexical learning mechanisms continue to be available in second language phoneme category formation.

Peperkamp et al. (2003, Experiment 2) investigated the role of phonological distributional information (i.e. whether phones occurred in overlapping or complementary distribution), in addition to statistical distributional information (i.e. whether phones occurred in a monomodal or bimodal distribution along some phonetic dimension), on the acquisition of phonemic and allophonic categories. Three groups of adult native French speakers received exposure to an artificially manipulated continuum of the voiced and voiceless uvular fricative [ʁ-χ], a French allophonic alternation, that varied with respect to both the statistical distribution and the phonological distribution. Unexpectedly, the interaction between exposure type and the difference in error rates on an AX discrimination task at the pre- and post-test was not significant. However, only the group exposed to the continuum in bimodal overlapping distribution showed a significant difference between pre-and post-test. The authors interpreted these results as being in line with their hypothesis that category formation is informed by both statistical information and information about the contexts in which individual segments occur.

However, the lack of a significant interaction makes it difficult to draw a definitive conclusion. The surprising results were attributed to a difference between the groups at the pretest, but other factors may have also reduced the likelihood of observing an effect, such as the variability in exposure tokens (i.e. 40 different words per target sound). Additionally, the phones making up the target contrast were related as allophones of the same phoneme in the participants' native language, a factor known to reduce listener sensitivity to a phonetic contrast (Boomershine, Hall, Hume, & Johnson, 2008). Thus, it may be that the exposure phase in the experiment could not override the lifetime of exposure to the alternation that the participants already had, or that any learning that could occur happened during the pre-test.

In another non-lexical perception study (i.e. one that does not require lexical access) investigating the acquisition of allophones by adult learners, Shea and Curtin (2010) asked whether learners are able to acquire knowledge of the relationship between L2 allophones and their phonological environments as they gain experience with the language. They compared four groups of listeners (monolingual English speakers, monolingual Spanish speakers, low-intermediate Spanish learners, high-intermediate Spanish learners) on a stress-detection task to determine whether more advanced learners performed more like native Spanish speakers with regard to expectations about the conditioning context of a Spanish allophonic alternant. In Spanish, voiced stops [b, d, g] alternate with approximants [β, ð, ɣ]; stops occur initially or after a nasal in the onset of a stressed syllable, whereas approximants occur intervocalically and are post-tonic. Like the native Spanish speakers, higher-proficiency Spanish learners showed a preference for associating the stressed syllable with initial stop consonants. Lower-proficiency learners and monolingual English speakers did not show this preference. The authors concluded that increasing experience with a second language results in increasing sensitivity to the factors that condition allophonic alternants, and they attributed this learning to a phonological distributional learning mechanism. However, the learning of this association occurred prior to the experiment, failing to clarify which aspects of their input contributed to this acquisition.

Taken together, although it is compelling to claim that a phonological distributional mechanism remains available for L2 phonological acquisition, no study to date has provided clear evidence that this mechanism is still available for use in adult allophonic acquisition. We attempt to build on this early literature by relying on participants with relatively little experience, controlling their input, and isolating phonological distributional information as an independent variable.

THE PRESENT STUDY

We investigate whether adult second language learners can use phonological distributional information in the absence of lexical-semantic information or explicit instruction to determine whether two phones belong to the same or different underlying categories. To this end, we expose relatively inexperienced listeners to a novel contrast, such that the two target phones occur in either overlapping or complementary distribution. We predict that learners trained on the contrast in overlapping distribution will infer two categories (a phonemic distinction), whereas learners trained on the contrast in complementary distribution will infer that the two phones are related as allophones of a single underlying phoneme category. Perceptually, adult native speakers are less sensitive to the phonetic distinction between allophones of the same phoneme (Whalen, Best, & Irwin, 1997; Boomershine, et al., 2008) than to allophones of separate phoneme categories, so we expect participants in the complementary exposure group category to perform less accurately on an ABX task than those in the overlapping exposure group. The ABX task was selected because it can be completed by participants who are naïve with respect to the tested language, yet it is typically assumed to tap into a more phonological level of representation than does the AX task used by Peperkamp et al. (2007).

Participants

Participants were 40 native English speakers (M = 12 F = 27 Other gender = 1) between the age of 17-52 years old (mean = 24.15) recruited at the University of Utah from the Salt Lake City area. Participants had studied Spanish for no more than one year, and they had no history of hearing, developmental, or neurological disorders. They were randomly assigned to one of two exposure groups: Complementary group (n = 20, M = 8 F = 11 Other gender = 1, mean age = 23.5) or Overlapping group (n = 20, M = 4 F = 16, mean age = 24.6). Participants were either compensated $10 for their time or received course credit for their participation.

Stimuli

The stimuli were natural recordings produced by a Spanish-English bilingual with phonetic training. Each non-word stimulus (see Figure 1) was produced in the carrier phrase *Diga la/una __________, por favor (*Say the/a _________, please*).* An acoustic analysis of the intensity ratios, a measure of degree of consonant constriction, was made for each production of the target phones [b] and [β] following Carrasco, Hualde, and Simonet (2012) to ensure there was indeed an acoustic distinction to learn (mean [β] intensity ratio = 82.26, mean [b] intensity ratio = 60.08, consistent with greater constriction for [b] than [β]). Four tokens of each stimulus type were selected for the task. The target sounds occurred in a syllable with [a] and the target syllable was either preceded or followed by [ti] and [ku]. While this is much less variability than typical of the real world, this was a conscious choice to reduce variability in other syllable types, given that Peperkamp et al. (2003) suggested the variability of their exposure tokens may have introduced unwanted noise in their study. Stress was controlled such that all stimuli had penultimate stress. The overlapping exposure condition contained [b] and [β] in both initial and medial position, whereas in the complementary condition [b] was restricted to initial position and [β] to medial position. Because the complementary exposure tokens were a subset of those used for overlapping exposure, each of the target nonwords in the complementary exposure condition was played twice as frequently to ensure both groups receive exposure phases of equal length. Control stimuli were identical to the target nonwords except the target contrast was replaced with a native contrast, [m-l]. The distribution of [m] and [l] was not constrained for either exposure group; both groups were exposed to tokens containing [m] and [l] in both initial and medial position. A summary of the stimuli is provided in Figure 1.

Complementary

		initial	medial
Experimental contrast	[b]	bati	
		baku	
	[β]		tiβa
			kuβa
Control contrast	[m]	mati	tima
		maku	kuma
	[l]	lati	tila
		laku	kula

Overlapping

		initial	medial
Experimental contrast	[b]	bati	tiba
		baku	kuba
	[β]	βati	tiβa
		βaku	kuβa
Control contrast	[m]	mati	tima
		maku	kuma
	[l]	lati	tila
		laku	kula

Figure 1. Stimuli heard in each exposure group.

Procedure

The experiment consisted of two parts: an exposure phase and a test phase.

Exposure phase

During the exposure phase participants heard 768 auditory tokens (12 randomized blocks x 4 tokens x 16 words) separated by an ISI of 500 ms. Participants were instructed to listen and check a box on the sheet provided each time they heard a word. The exposure lasted approximately 16 minutes.

Test phase

The test phase consisted of an ABX discrimination task in which the participant indicated whether the third sound was the same as the first or the second. Each token (4 tokens x 16 words = 64 trials) appeared as X; in each trial, the three words were all physically different tokens. Trials were counterbalanced for A and B matches and were presented in a random order for each participant. The test was divided into two blocks of 32 trials. Test items were separated by an ISI of 500 ms and trials timed out at 2500 ms.[1]

RESULTS

Two analyses were conducted. For the first, the dependent measure was mean proportion correct, and for the second, the dependent measure was d-prime score.

[1] Although the complementary distribution group was exposed to only [b] initially and [β] medially during the exposure phase, due to the nature of an ABX task, during the test phase they received input to each target token in the “wrong” position. For example, in trial [bati] – [βati] – [βati], participants were exposed to two instances of initial [β]. We ensured that the ratio of “wrong” test items to “right” training items remained low (9%). This concern informed the duration of the exposure phase as well as the number of trials in the test phase.

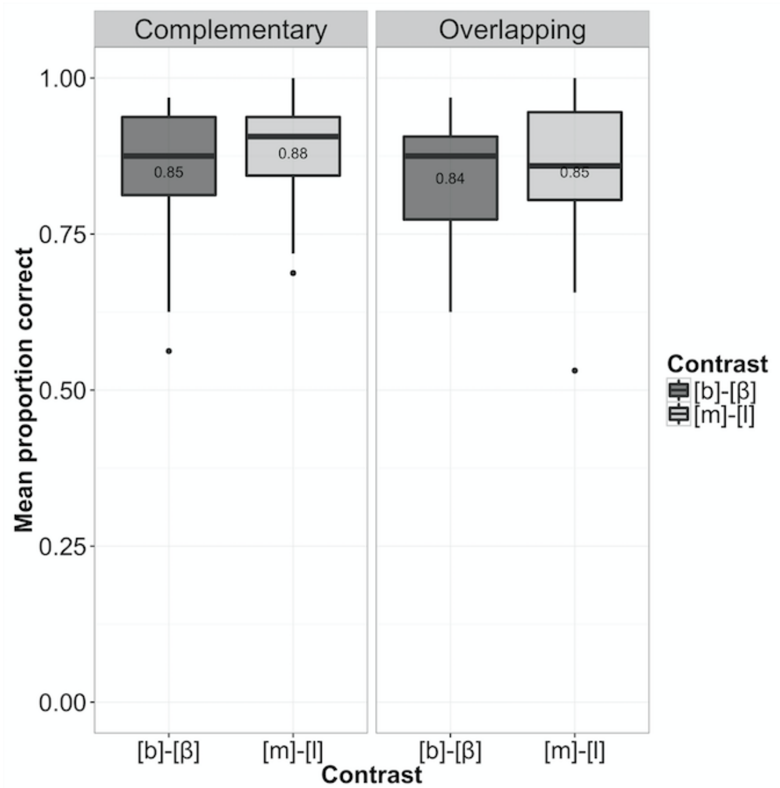


Figure 2. Boxplot of Mean Proportion Correct by Exposure Group and Contrast. Superimposed text represents group means.

As seen in Figure 2, the mean proportion correct for the complementary exposure group was 0.85 for the target contrast and 0.88 for the control contrast. For the overlapping exposure group, mean proportion correct for the target contrast was 0.84 and for the control contrast was 0.85. A two-factor mixed-design ANOVA was conducted with exposure group as the between-subjects factor (two levels: complementary and overlapping) and contrast as the within-subjects factor (two levels: [b- β] and [m-l]). There was no significant main effect of exposure group $F(1, 38) = 0.498$, $p = 0.484$ or of contrast $F(1, 38) = 2.791$, $p = 0.103$. Crucially, there was no significant interaction between group and contrast $F(1, 38) = 0.112$, $p = 0.740$, indicating that the two groups did not differ from one another in their performance on the [b-β] contrast.

In addition to considering the mean proportion correct, d' scores were also computed and submitted to statistical analysis. D-prime is a measure of sensitivity that factors out individual response bias.

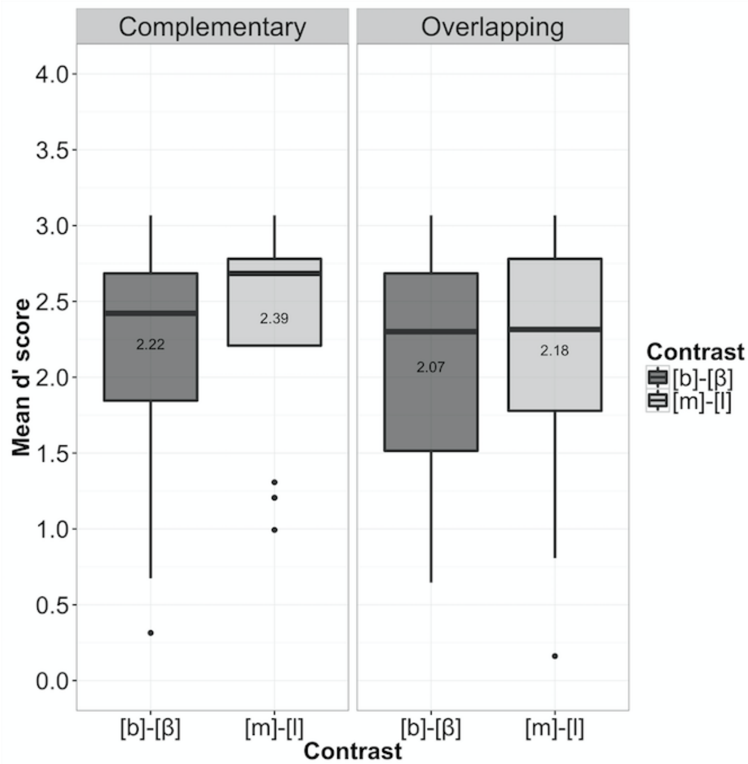


Figure 3. Boxplot of D-Prime Scores by Exposure Group and Contrast. Superimposed text represent group means.

As seen in Figure 3, the mean d' score for the complementary group was 2.22 for the target contrast and 2.39 for the control contrast. For the overlapping group, mean d' score for the target contrast was 2.07 and for the control contrast was 2.18. A second two-factor mixed-design ANOVA was performed. Again, there was no significant main effect of exposure group $F(1, 38) = 0.698$, $p = 0.409$ or of contrast $F(1, 38) = 2.754$, $p = 0.105$, and there was no interaction of group and contrast $F(1, 38) = 0.140$, $p = 0.710$.

DISCUSSION

This study investigated the role of phonological distributional information in adult L2 phonological acquisition. We asked whether participants exposed to a novel phonetic contrast [b-β] in either complementary or overlapping phonological distribution performed differentially on an ABX task following the exposure. Unexpectedly, we observed no difference in the two groups' performance on the [b-β] contrast, suggesting that participants were not able to use phonological distributional information alone to determine whether the two phones were allophones of the same phoneme or whether they were different phonemes. These results are consistent with those of Peperkamp et al. (2003), suggesting that perhaps, at least under these constrained laboratory conditions, a phonological distributional mechanism alone may be insufficient for adult learners to infer the phonological status of a novel pair of phones.

We attempted to build on earlier work by controlling for previous input regarding the novel phones, using a measure of perception that relies on higher levels of phonological representation, and isolating the variable of phonological distributional information as a potential contributor to the acquisition of a novel allophonic alternation. Previous research (i.e. Peperkamp et al., 2003; Shea & Curtin, 2010) used stimuli created by splicing allophones produced in their conditioned contexts into phonotactically illegal contexts, while the stimuli in this study were naturally produced in illegal contexts. This solution may have eliminated extraneous cues introduced by cross-splicing. However, it also had limitations, such as the difficulty of producing phones in unnatural positions and the necessity of ensuring that productions in illegal contexts were produced with similar intensity ratios as those in legal contexts. The stimuli were chosen based on a phonetic analysis of the intensity ratio between the target consonant and the following vowel [a] as well as the presence/absence of a burst. Carrasco et al. (2012) showed that there is a bimodal distribution of the intensity ratios of [b] and [β] in both Madrid and Costa Rican Spanish, so we sought to verify that a similar distribution was present in our stimuli. However, some research suggests that allophones are not only perceived as more similar than contrastive phones, but that they are also produced more similarly in languages for which they are allophonic than they are in languages for which they are phonemic (see e.g., Seidl & Cristia, 2012). It is possible that we artificially enhanced the statistical distribution of these phones due to this analysis.

The task used in this experiment was carefully considered, as the foundational literature in L2 allophone acquisition is relatively sparse. Shea and Curtin (2010) successfully showed gains in allophonic knowledge of learners using a metalinguistic task that demonstrated understanding of allophones with respect to their conditioning contexts. Because we asked a different question—whether learners could make use of phonological distributional information to group phones as allophones of a single phoneme category—we chose a different task. Using a task that relied on the perceptual correlates of phonemes and allophones as a measure was a less metalinguistic way of detecting early knowledge about phonological status. This task also had limitations, such as the necessity of including tokens that one of the groups did not hear during the exposure phase in the test and the possibility that phonological status could be "unlearned" through this conflicting evidence. It is also possible that participants used an acoustic strategy, such as treating B and X as an AX task and ignoring A altogether, a possibility that could be effectively eliminated by using more than one talker to produce stimuli in future research.

At the same time, the null effects may have been caused by an unexpected pattern of perceptual assimilation. The alternation was chosen to be maximally similar to Shea and Curtin (2010), who looked at the entire natural class that undergoes this alternation in Spanish, /b,d,g/. We considered the alternation [g]-[ɣ] because it would likely result in a same-category assimilation in English (Best, 1995), but we were concerned about the perceptual salience of this distinction. Additionally, an acoustic analysis of intensity ratios found that these phones are monomodally, as opposed to bimodally, distributed in Spanish (Carrasco et al., 2012). We avoided the third pair [d]-[ð] because they are contrastive in English. English does not have a category /β/, so the study was designed as if participants would hear one familiar sound /b/ and another that was also assimilated to /b/ but may have been a poorer exemplar of the category. However, pilot studies indicated that [β] was frequently assimilated to the category /v/ by English speakers. It could be that this two-category assimilation (Best, 1995) blocked distributional learning during the short exposure period. While we strategically chose a pair that contained one familiar and one unfamiliar phone for English speakers, it is worth noting that the L1 English L2 learner of Spanish would be required to master all of these assimilation patterns.

Moreover, although earlier studies have shown that infants can learn from distributional information with as few as two minutes of exposure (Saffran et al., 1996) and adults with as few as nine (Hayes-Harb, 2007; Peperkamp et al., 2003), it is also possible that a longer exposure period may be required to learn this particular alternation or allophonic alternations in general.

Finally, we would have ideally used participants who were completely naïve to Spanish, but it proved challenging to find participants without Spanish experience in the Salt Lake City area. It is possible that informal and/or limited formal exposure to Spanish (which was not controlled for in this study) provided prior input from which participants could have learned about the alternation.

This study attempted to tightly control many variables that researchers noted may have contributed statistical noise in their studies and masked small but reliable effects from being detected. Despite these modifications, we failed to find support for the hypothesis that adult second language learners are able to use phonological distributional information alone to learn that two phones are related as allophones of the same phoneme. Nevertheless, there is much to build upon. For future research, other ways of selecting stimulus tokens should be considered and other contrasts should be investigated. Alternations from other languages may be better when a large number of participants in an area speaks Spanish as an L2. Despite the null results, the role of phonological distributional information is an important area of research in which study should continue in order to deepen our understanding of adult second language phonological acquisition.

ACKNOWLEDGMENTS

The authors are grateful to members of the Speech Acquisition Lab at the University of Utah for their contributions to this work, as well as the audiences at the 2016 Pronunciation in Second Language Learning and Teaching Conference in Calgary, Alberta for their comments and feedback.

ABOUT THE AUTHORS

Taylor Anne Barriuso (tayloranne.barriuso@utah.edu) is a PhD student in the Department of Linguistics at the University of Utah. She has studied the influence of grammaticality and instructor language background in L2 pronunciation in addition to features of the input available for learning novel allophonic alternations. She is currently investigating whether perceptual gains from high variability phonetic training can be exploited to establish and use lexical representations that encode difficult L2 phonological contrasts.

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 range of experimental techniques to examine the role of existing phonological knowledge in the representation and processing of nonnative input, with the goal of better understanding the nature and persistence of cross-language influence in speech perception and the acquisition of an L2 sound system.

Contact Information for both authors: Department of Linguistics, Languages and Communications Building, University of Utah, 255 South Central Campus Drive, Salt Lake City, UT 84112-0942, USA

REFERENCES

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.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M.J. Munro & O.-S. Bohn (Eds.), *Second language speech learning: The role of language experience in speech perception and production* (pp. 13-34). Amsterdam: John Benjamins.

Boomershine, A., Hall, K. C., Hume, E., & Johnson, K. (2008). The impact of allophony versus contrast on speech perception. In P. Avery, B. Dresher & K. Rice (Eds.), *Contrasts in phonology: Theory, perception, acquisition* (pp. 145-171). Berlin: Mouton de Gruyter.

Carrasco, P., Hualde, J. I., & Simonet, M. (2012). Dialectal differences in Spanish voiced obstruent allophony: Costa Rican versus Iberian Spanish. *Phonetica*, *69*(3), 149-179.

Chambers, K. E., Onishi, K. H., & Fisher, C. (2003). Infants learn phonotactic regularities from brief auditory experience. *Cognition*, *87*(2), B69-B77.

Eckman, F. R. (1981). On the naturalness of interlanguage phonological rules. *Language Learning*, *31*(1), 195-216.

Feldman, N. H., Myers, E. B., White, K. S., Griffiths, T. L., & Morgan, J. L. (2013). Word-level information influences phonetic learning in adults and infants. *Cognition*, *127*(3), 427-438.

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.

Hayes-Harb, R. (2007). Lexical and statistical evidence in the acquisition of second language phonemes. *Second Language Research*, *23*(1), 65-94.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*(5044), 606-608.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), B101-B111.

Peperkamp, S., Le Calvez, R., Nadal, J. P., & Dupoux, E. (2006). The acquisition of allophonic rules: Statistical learning with linguistic constraints. *Cognition*, *101*(3), B31-B41.

Peperkamp, S., Pettinato, M., & Dupoux, E. (2003). Allophonic variation and the acquisition of phoneme categories. In B. Beachley, A. Brown, & F. Conlin (Eds.), *Proceedings of the 27th Annual Boston University Conference on Language Development,* (pp. 650-661). Sommerville, MA: Cascadilla Press.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926-1928.

Seidl, A., & Cristia, A. (2012) Infants' learning of phonological status. *Frontiers in Psychology, 3(448)*, 1-10.

Shea, C. E., & Curtin, S. (2010). Discovering the relationship between context and allophones in a second language. *Studies in Second Language Acquisition*, *32*(4), 581-606.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*(1), 49-63.

Whalen, D. H., Best, C. T., & Irwin, J. R. (1997). Lexical effects in the perception and production of American English /p/ allophones. *Journal of Phonetics*, *25*(4), 501-528.

White, K. S., Peperkamp, S., Kirk, C., & Morgan, J. L. (2008). Rapid acquisition of phonological alternations by infants. *Cognition*, *107*(1), 238-265.

Yeung, H. H., & Werker, J. F. (2009). Learning words' sounds before learning how words sound: 9-month-olds use distinct objects as cues to categorize speech information. *Cognition*, *113*(2), 234-243.

Becker, S. (2017). Perception and production of unfamiliar L2 segments: Using technology for teaching and research. 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. 21-32). Ames, IA: Iowa State University.

PERCEPTION AND PRODCUTION OF UNFAMILIAR L2 SEGMENTS: USING TECHNOLOGY FOR TEACHING AND RESEARCH

Shannon Becker, Northern Illinois University

In this paper I discuss the first step in an ongoing empirical study on the perception and production of French nasal vowels in a classroom context. The first phase of the study, presented here, analyzes L1 English learners' perception and production of L2 French nasal vowels /ɑ̃/ and /ɔ̃/, an important contrast in spoken French, as it distinguishes between many minimal pairs in the lexicon as well as between gerundive and nominal forms of verbs. Since nasality is not phonemic in English, and because both L2 segments are open back vowels differing mostly in roundedness, learners have difficulty perceiving the difference between these segments. This leads to trouble pronouncing them as distinct phones in spontaneous production. Results from the pilot study reveal that, regardless of level, university-level French learners are not improving with regard to the target sounds in either perception or production.

In addition to presenting these findings, *Audacity* audio recording software, *Praat* speech analysis software, and *Qualtrics* survey software are discussed as a means of incorporating listening and pronunciation practice into classroom language learning in order to draw learners' attention to important phonetic features of the language and to improve their perception and pronunciation.

INTRODUCTION

While production is generally regarded as the hallmark of proficiency in second language performance, a question that is often overlooked in practice, if not in research, is whether the L2 speaker is correctly perceiving the sound she wishes to produce. If one cannot accurately perceive the acoustic-phonetic and articulatory properties of a sound, it may be difficult to accurately produce that sound. This is especially true in the case of sounds that occur in allophonic variation in the L1 but are phonemic in the L2, as in the case of native speakers of American English (AE) learning the four nasal vowels that are essential to the accurate pronunciation of L2 French.

Nasal vowels represent a possible area of difficulty for English-speaking learners of L2 French due to their lack of phonemic distinction in the L1. It can be difficult for L1 AE learners of French to perceive the contrast between oral and nasal vowels as well as the contrasts among the various nasal vowels, and this, in turn, may be related to difficulty producing nasal vowels in the appropriate phonetic environments. In this paper, I describe the results of a pilot study focused on L1 AE learners' perception and production of the French nasal vowels /ɑ̃/ and /ɔ̃/; how the acquisition of these phones is affected by typical classroom L2 instruction; and how the technological tools that were used in this research study can be implemented in the classroom in order to help learners improve their performance on nasal vowel contrasts.

REVIEW OF PREVIOUS LITERATURE

Perception and Production of French Nasal Vowels

Two issues are addressed in the current study: the relationship between perception and production of L2 segments and the ability of second language learners to improve their perception and production of French nasal vowels. Perception of segments in an L2 can have important effects on the speed of language processing (Munro & Derwing, 1995), the ease of word recognition (Bradlow & Pisoni, 1999), and the production of those segments (Rochet, 1995). There is no doubt that a link exists between perception and production; the nature of that link is, however, unclear. Perception is understood by some researchers to precede production in L2 acquisition (Flege, 1995), while others posit that the two processes occur in parallel (Best, 1995), and still others have provided evidence that production can in fact precede perception (Sheldon & Strange, 1982). The relationship between perception and production, and how these two skills interact in the language acquisition process, is an essential aspect of L2 proficiency.

As Zampini (2008) has observed, there is a notable lack of research into the production of L2 nasal vowels. Similarly, there appear to have been relatively few studies on the perception of nasal vowels among L2 learners. Where perception studies have been performed with oral vowels, the results have often differed considerably from those obtained from consonant perception studies. While scholars have generally found that improved perception resulting from the training of certain consonant contrasts can be transferred to untrained contrasts, perception of vowel contrasts has not proven as transferable (Nishi & Kewley-Port, 2008, p. 1480).

Research on the perception and production of French vowels has focused almost exclusively on oral vowel contrasts. For example, Levy and Law (2010) looked at the effects of language experience and consonantal context on the production of French /i – y – u – œ – a/ in bilabial and alveolar contexts. Rochet (1995) evaluated acquisition of the French /y/ by speakers of languages with only /i/ and /u/, both in terms of current ability and the potential of perceptual training. Perceptual training studies have also focused on important oral contrasts such as /y – u/ (Simon, Chambless & Ubirata Kickhoefel, 2010) and /ə – o/ (Brosseau-Lapre, Rvachew, Clayards, & Dixon, 2013). A notable exception is a study by Inceoglu (2016), which found that audio-only and audiovisual training both significantly improved L1 AE learners' perception and production of Parisian French nasal vowels /ɔ̃ – ɑ̃ – ɛ̃/ compared to a control group.

More research is necessary to identify areas of difficulty for L1 AE learners of French nasal vowels, specifically which contrasts are least easily acquired and how perception and production interact in this acquisition.

French Nasal Vowel Pair /ɑ̃/ – /ɔ̃/

French contains four nasal vowels: /œ̃/, /ɔ̃/, /ɛ̃/, and /ɑ̃/. The present study focuses on non-native listeners' perception of /ɑ̃/ and /ɔ̃/. These vowels are important in French because they represent a strong phonemic contrast. For example, the words *sans* /sɑ̃/ and *son* /sɔ̃/ (*without* and *his/her*, respectively) constitute a minimal pair with important semantic implications. Additionally, students demonstrate difficulty in perceiving these sounds correctly and producing them accurately. Finally, the oral counterparts of these vowels exist in the L1 English inventory while the nasal versions exist only in allophonic variation.

Results from an MRI study with cepstral analysis by Delvaux, Metens, and Soquet (2002) showed that /ɑ̃/ is articulated somewhat lower, more rounded, and more posterior than its oral counterpart /a/. An analysis of the /ɔ – ɔ̃/ pair revealed that the nasal vowel /ɔ̃/ is more rounded than its oral counterpart. The acoustic analysis for both pairs demonstrated a drop in F2 that corresponds to the shifting of the velum for nasal vowels. French nasal vowels thus demonstrate important acoustic-phonetic differences from their oral counterparts.

Following Flege's Speech Learning Model (SLM) (1995), it is likely that L1 English learners of French classify the French oral vowels /a/ and /ɔ/ as phones identical to those in their existing American English inventory. I propose that the French vowels /ɑ̃/ and /ɔ̃/ are categorized as similar L2 phones, making them more difficult to distinguish. First, since the oral vowels /ɑ/ and /ɔ/ exist in both the L1 and L2 inventories and are realized in phonetically similar ways, the articulatory and phonetic changes necessary to nasalize them may not represent a distinction sufficient for L2 learners to create a 'new' phonetic category.

It is also common in American English for vowels to be nasalized due to the phonetic environment or to individual differences in the realization of the phones. Oral and nasalized vowels in L1 AE do not represent a phonemic distinction, so these phonetic properties may not be salient to L1 AE speakers when listening in L2 French. Finally, the articulatory and phonetic changes that occur in the shift from oral to nasal vowels in French may mute the distinction between the two vowels for AE listeners. In other words, the perceptual distinction that exists between the oral vowels /ɑ/ and /ɔ/ may be lost in the movement, rounding, and drop in F2 values that occur when these vowels become nasal, leading to an inability to perceive the difference between /ɑ̃/ and /ɔ̃/.

Research Questions

1. Given their current instruction, are students progressing in their perception and production of French nasal vowels, specifically the /ɑ̃ – ɔ̃/ contrast?
2. Is there a demonstrable relationship between students' ability to discriminate between these sounds and their ability to produce them?

METHOD

The pilot study presented here was undertaken as the first step in a research project that will use high variability phonetic training (HVPT) (Logan, Lively, & Pisoni, 1991) to aid L1 English learners of L2 French in the development of the necessary perceptual skills to distinguish French nasal vowels. To establish the need for such a program, the pilot study assessed learners' acquisition of French nasal vowels at the beginning and end of a semester of typical instruction, with no experimental intervention.

Participants

The participants in the study were 33 L1 American English learners of French, ranging in age from 18 to 35 with an average age of 22. All reported normal hearing. In order to look at differences according to level of instruction, participants were recruited from three levels: 102 (second semester), n = 13; 202 (fourth semester), n = 15; and 302 (sixth semester), n = 5. One female native speaker of French served as the speaker for the perceptual tasks.

Materials

The perceptual stimuli consisted of 40 one- and two-syllable word pairs, embedded in individual words (30) and short sentences (10). Each word pair contained either the target contrast (/ɑ̃ – ɔ̃/) or a distractor contrast (/e – i/, for example). There was a total of 17 tokens of the target vowel contrast, 12 in the list of words and 5 in the list of sentences. For the production task, one word was chosen at random from the word pairs used in the perceptual task to create a list of 30 words and 10 short sentences to record. Participants were asked to perceive and produce words containing nasal vowels both in individual words and in short sentences in order to determine if the additional context of the phonetic environment in a sentence had an impact on their ability to discriminate and produce sounds. An example of a word pair used in the perceptual task is <violon> (/vjɔ.lɔ̃/, *violin*), which was contrasted with <violent> (/vjɔ.lɑ̃/, *violent*), a pair with important semantic implications.

Procedure

After the initial recruitment, participants scheduled two separate appointments. During the first appointment, after filling out a demographic questionnaire, they were sent a link to a *Qualtrics* survey containing a forced choice perception task in which they heard the series of randomized word pairs followed by the series of randomized short sentences. In each series, they were asked to indicate the word they heard from two choices. They were then provided the list of words and short sentences and asked to record these in two *Audacity* files, which were then analyzed using *Praat* speech analysis software. This procedure, with the exception of the demographic questionnaire, was also followed at the second appointment, which took place approximately ten weeks later.

RESULTS

Perception

Results from the perceptual task were inputted into IBM *SPSS Statistics* and were analyzed to determine whether there was a significant improvement from pretest to posttest on either word pairs or sentences at each level; how the different levels compared to one another at each time step (pretest and posttest); and whether improvement over time was significantly different among the class levels.

A series of paired-samples t-tests determined that the level of improvement from pretest to posttest was not significant for any class level in either phonetic context (see Table 1). A repeated measures analysis of variance confirmed that there were no significant differences based on class level for either words (p = .553) or sentences (p = .733). Similarly, there were no significant differences between groups with regards to improvement over time (posttest score – pretest score) for either words (p = .299) or sentences (p = .196).

Table 1

Paired-samples t-test results comparing perceptual results by class level over time (p-values)

Level	Words				Sentences			
	t	*M*	*SD*	*p*	*t*	*M*	*SD*	*p*
102	.65(12)	2.56	14.19	.527	1.08(12)	6.15	20.63	.303
202	1.78(14)	7.22	15.71	.097	-1.57(14)	-8.00	19.71	.138
302	-0.69(4)	-5.00	16.24	.529	0.34(4)	4.00	26.08	.749

Additionally, despite the lack of statistical significance among levels, the general trend shows that participants perceived vowel differences better overall (i.e., in all combined vowel combinations) than they did for the target nasal vowels as a subset. This was true at both the pretest and the posttest (see Tables 2 and 3). While students in second-semester French (102) did not see an improvement in nasal vowel distinctions in any context, fourth-semester (202) learners showed considerable improvement in target nasal vowel contrasts embedded in short sentences, while sixth-semester (302) students seemed to improve in target contrasts embedded in words. Only the sixth-semester students showed any improvement in the combined score.

Table 2 presents the pretest and posttest scores, along with improvement over time, on the perceptual task for vowel contrasts embedded in word pairs. "Target" scores represent performance only on the target vowel contrasts, while "Overall" scores represent all combinations including distractors (oral vowel contrasts as well as oral-nasal contrasts). Scores are represented here as the percentage of correct responses.

Table 2

Average performance on target contrasts embedded in words

Level	Pretest		Posttest		Improvement	
	Target	*Overall*	*Target*	*Overall*	*Target*	*Overall*
102	67.31	73.59	64.74	74.1	-2.57	0.51
202	76.11	80	68.89	76	-7.22	-4
302	66.67	76.67	71.67	80	5	3.33

Table 3 shows the pretest and posttest scores, along with improvement over time, on the perceptual task for vowel contrasts embedded in short sentences.

Table 3

Average performance on target contrasts embedded in short sentences

Level	Pretest		Posttest		Improvement	
	Target	Overall	Target	Overall	Target	Overall
102	69.23	76.92	63.08	72.31	-6.15	-4.61
202	66.67	77.33	74.67	80	8	2.67
302	68	76	64	72	-4	-4

Production

As mentioned earlier, Delvaux et al. (2002) have identified four elements that distinguish French nasal vowels from their oral counterparts. In terms of articulation, they are lower, more rounded, and more posterior. Acoustically, they tend to have a lower F2 value, an indication that the velum has lowered to allow air to pass through the nasal cavity. Since the current study did not make use of the MRI, in this section I will describe some general tendencies in the production of the vowels from participants in each class level.

To illustrate the differences between word-initial and word-final nasal vowels, I have chosen the words <angle> /ɑ̃gl/, <quand> /kɑ̃/, <onze> /ɔ̃z/, and <blond> /blɔ̃/. I carefully analyzed and transcribed the students' productions of these words and calculated the rate of correct responses on the basis of these analyses. The results are presented in Table 4 below. The scores represent the number of students at each class level who produced the nasal vowel correctly, irrespective of how they pronounced the other sounds in the word. The only exception to this rule is for those who produced the nasal consonant /n/ after the nasal vowel, since the addition of the consonant reduces the nasality of the vowel.

Table 4

Number of correct nasal vowel productions by class level

Level	angle	quand	onze	blond
102	9/13	5/13	7/13	11/13
202	11/15	6/15	5/15	10/15
302	2/5	3/5	2/5	3/5

Word-initial /ɑ̃/. While <angle> presented some problems for students at all class levels, the /ɑ̃ – ɔ̃/ distinction was not entirely to blame. One learner each in 102 and 302 produced a sound that fell between /ɑ̃/ and /ɔ̃/, and two learners each in 102, 202, and 302 produced the nasal consonant /n/ after the initial vowel, thereby reducing the vowel's nasalization. Two errors occurred in non-target sounds, with the variants [ɑ̃gle] and [ɑ̃glo] being produced by 102 learners. Only two of the 302 students produced the nasal vowel correctly for this word; however, two of the incorrect responses were due to the addition of /n/ after the vowel. Since only five students from this level were included in the study, the percentages were easily reduced by these minor errors.

Word-final /ɑ̃/. <Quand> was the most difficult of the four words in terms of the variety of incorrect sounds produced. As expected, students in 102 experienced more difficulty than those in later stages of instruction, producing variants such as [kwan], [kɑ̃n], and [kɔ̃]. Seven of the 13 learners in this group produced a vowel sound other than /ɑ̃/, including /a/, /æ̃/, /ɔ/, and /ɔ̃/. Five of these learners added the /n/ sound after the vowel. Similarly, students from level 202 pronounced this word as [kɑ̃n], [kɑ̃nd], and [kɑ̃d], among others. In this group, six of the 15 learners produced a sound other than /ɑ̃/. In contrast to the 102 group, however, all 202 learners who used the wrong vowel produced it as /ɔ̃/. Three students in this group added either /n/ or /ŋ/ after the vowel. Finally, although this word provoked difficulty for the 302 students, their errors involved only the amount of rounding in the nasal vowel rather than incorrectly inserting consonants or using oral vowels like the learners from the lower levels did. That is to say, two of the 302 students produced the word as [kɔ̃] and the other three correctly pronounced [kɑ̃].

Word-initial /ɔ̃/. The word <onze> generated only minor errors for 102 and 202 students. In 102, the six learners who were marked wrong pronounced the nasal consonant /n/ after the nasal vowel: [ɔ̃nz]. No one in the group produced the vowel as the unrounded /ɑ̃/. In 202, of the 10 learners who were marked wrong, nine of them added the nasal consonant /n/. Only one 202 learner produced a vowel sound that fell between /ɑ̃/ and /ɔ̃/. Three students in 302 added /n/ after the nasal vowel; however, all of them produced the vowel itself correctly. That is to say, they were seemingly aware of using the necessary rounding, but failed to leave out the nasal consonant.

Word-final /ɔ̃/. <Blond> seems to have been the easiest of the four words to produce for 102 students. Eleven out of 13 produced it correctly, with only one student producing a nasal sound somewhere between /ɑ̃/ and /ɔ̃/ and one student adding /n/. Ten of the 15 students from 202 produced this sound correctly; of the five who did not, four pronounced the /ɔ̃/ as / ɑ̃/ and one added /n/. Three 302 students produced the sound correctly and the other two produced a sound in between /ɑ̃/ and /ɔ̃/. That is to say, while they may have been aware of the difference between the two sounds, they failed to fully round the /ɔ̃/.

Interestingly, the trends in perception do not line up with the trends in production for all words at all levels. For example, while <quand> appears to have been the most difficult word for 202 learners to produce, with six of 15 producing it accurately, the perception data suggest that it was among the easiest for them to perceive, with 11 of 15 participants correctly identifying it. Conversely, although <blond> seems to have been the easiest word for 102 learners to produce, with 11 of 13 students producing it accurately, it actually received the lowest scores on perception, with only seven of 13 students in 102 correctly identifying it. Despite the inconsistent, relatively inaccurate production performance of the 302 group, they performed better on the perception task, with four out of five students correctly perceiving all four words analyzed here. These results indicate a disconnect between perception and production for learners at all class levels regarding nasal vowels.

DISCUSSION

Perception

The results of the perceptual tasks demonstrate empirically what many French instructors have observed anecdotally in their classrooms: that students at all levels of instruction have a hard time distinguishing among the various French nasal vowels, in this case between /ɑ̃/ and /ɔ̃/. As expected, there were no statistically significant differences between groups or over time, despite the appearance of some divergent performance. This supports the hypothesis that learners are not progressing in their perception of nonnative L2 segments over the course of their typical classroom instruction.

Of note, however, are the results demonstrating that 202 students improved their perception in the context of short sentences while 302 students improved in distinguishing target contrasts in individual words. This could be because short sentences provide extra contextual and phonetic information whereas single words do not, potentially explaining why the advanced students improved in the more difficult context of single words while intermediate students required the extra contextual information provided by short sentences.

Despite some small differences among groups, the results of the perception tasks in this pilot study demonstrate that in the course of their normal instruction, learners are not advancing in their ability to perceive French nasal vowels. Future research should implement an instructional technique that provides learners with plentiful opportunities to hear French nasal vowels in varying phonetic contexts, and guides them in perceiving these differences. An improvement in perception compared to a control group may be accompanied by a parallel improvement in production of these segments.

Production

The production data presented here do not illustrate any consistent trends among levels or phonetic contexts. For example, while learners had difficulty with the word-final /ɑ̃/ in <quand>, they performed considerably better with word-final /ɔ̃/ in <blond>. This could be explained by first-language interference due to orthography, since <qu> in English is pronounced /kw/. However, whereas word-initial /ɑ̃/ in <angle> was produced correctly by most students, word-initial /ɔ̃/ in <onze> was frequently produced as [ɔ̃n].

The recordings used in this study provide general information about production trends based on class level as well as details of individual learners' productions. The inconsistent nature of these production results and their relative lack of correlation to performance on the perception task, provide evidence that sufficient progress is not being made in the case of nasal vowels.

Use of Technology in Teaching and Research

Many of the programs that we utilize as researchers can also be successfully implemented in L2 classroom instruction. Research has shown that metacognitive awareness can aid learners in recognizing weaknesses and improving the cognitive processes they use to process input (Vandergrift, 2002; Vandergrift & Tafaghodtari, 2010). To this end, survey software such as *Qualtrics* as well as audio recording programs such as *Audacity* and speech analysis software such as *Praat* can be leveraged to give students the practice and the metacognitive tools to work on their perception and production.

Listening and pronunciation practice are two exercises that are frequently lacking in L2 classroom instruction where time is at a premium and the focus is generally on communicative learning. This is where online survey software like *Qualtrics*[1] can be utilized in conjunction with course management software to assign these types of activities as homework. *Qualtrics* allows for certain functions that aren't available in most course management programs, such as the embedding of audio files into questions, randomization of items, automatic scoring, and even somewhat primitive reaction time measures. The instructor can record items using *Audacity* and upload them into the *Qualtrics* library to insert them into future surveys. Perception of words and segments can be verified via multiple choice, short answer, and essay questions. A link to the *Qualtrics* survey can easily be embedded in the course management program. While the process of uploading audio files entails an initial output of time, this is mitigated by the ability to reuse audio files from semester to semester. Over time, various speakers can be recorded in order to build a database of audio excerpts that exemplify both phonetic and speaker variability.

Audacity[2] is a user-friendly audio recording program that is free to download, making it incredibly useful for L2 instructors and students. As a pronunciation exercise, for example, an instructor might make a recording of a list of words and phrases produced by several native (or nativelike) speakers and upload it to the course management system. Students could then be instructed to listen and repeat as often as necessary until they feel they have mastered the pronunciation, and then to record themselves reading the list and post it privately in the course management system for the instructor to listen and provide feedback.

For more advanced students, speech analysis exercises can be useful in analyzing their own pronunciation and having a visual representation of their progress. Without broaching advanced linguistic concepts, they can learn how to import their audio files into *Praat*, view and edit spectrograms, and compare them visually to their own previous recordings of the same items. They can experiment with different pronunciations and see how these compare to each other in *Praat*. For visual and kinesthetic leaners, this type of exercise has the potential to be a powerful tool to help them grasp how subtle differences in pronunciation can change output (Lambacher, 1999; Saito, 2007)

These programs, in addition to being useful in language acquisition research, allow educators to more easily incorporate listening and pronunciation practice into their courses despite time constraints. As an added benefit, aligning the programs we use in empirical studies with what we use in the classroom represents a step toward bridging the gap between theory and practice.

Further Research

Nasal vowels represent an area of particular difficulty for L1 American English learners of French. In this preliminary study I have provided evidence of the need for targeted instruction in the perception and production of these sounds, a task that can be made easier through the implementation of accessible online resources such as those used in the study. Moving forward, I

[1] Institutional subscriptions range from $500-$1000: www.qualtrics.com
[2] Free to download for PC and MAC at www.audacityteam.org

will also test whether HVPT (Logan et al., 1991) has the potential to provide the necessary input and cognitive and metacognitive procedures to lead to successful perception and production of unfamiliar L2 segments such as French nasal vowels.

In recent years, there has been increased interest in the concept of training learners to perceive differences in L2 vowel sounds (Nishi & Kewley-Port, 2008; Tajima, Kato, Rothwell, Akahane-Yamada, & Munhall, 2007; Thomson, 2012). Thomson (2012), for example, in his training of 26 Mandarin L1 speakers to perceive English vowels, found that learners' perception of English vowels had improved in the training context as well as in one novel phonetic context.

I posit that training L1 AE speakers to perceive French nasal vowel contrasts can be achieved via HVPT because it encourages the use of stimuli presented in multiple phonetic contexts and by multiple native speakers to build robust perception of target contrasts. The perceptual difficulty L1 AE learners have in acquiring French nasal vowels makes them particularly good candidates for the use of this type of training. Through repeated exposure to the target sounds in various phonetic contexts, learners may become more familiar with the acoustic-phonetic properties of these phones and develop the perceptual processes to distinguish among them in novel contexts.

CONCLUSION

In this study I have presented empirical evidence from three levels of French instruction to demonstrate that, given the current instructional paradigm, students are not progressing in their development of perception and production skills with regard to French nasal vowels. At all levels of instruction, performance on a perceptual task did not improve over time. Students at higher levels did not perform significantly better than those at lower levels either at pretest or at posttest Production performance was inconsistent in terms of phonetic context as well as relationship to perception.

Because French nasal vowels are likely assimilated as similar phones for L1 AE speakers, and because they have an impact on meaning due to their phonemic nature in the L2, it is imperative that L2 French listeners learn to differentiate between them, both in perception and production. Previous research has shown that perceptual training can have an impact on vowel perception; however, nasal vowels have remained relatively unstudied. The results of this pilot study support the need for effective pronunciation instruction, which will be studied further using high variability phonetic training.

ABOUT THE AUTHOR

Shannon Becker is an Assistant Professor of French Linguistics at Northern Illinois University. Her research interests include development of L2 listening comprehension skills; learner perception and production of French segmentals; L2 pedagogy; complex adaptive systems/dynamic systems approaches to SLA; and technology for the foreign language classroom.

Watson Hall 115
Northern Illinois University
815-753-1501
sbecker@niu.edu

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. 171-204). Timonium, MD: York Press.

Bradlow, A. R. & Pisoni, D. B. (1999). Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors. *The Journal of the Acoustical Society of America, 106*(4), 2074-2085.

Brosseau-Lapre, F., Rvachew, S., Clayards, M., & Dickson, D. (2013). Stimulus variability and percpetual learning of nonnative vowel categories. *Applied Psycholinguistics, 34*(3), 419-441.

Delvaux, V., Metens, T., & Soquet, A. (2002). French nasal vowels: acoustic and articulatory properties. *Proc of the 7th ICSLP*, *1*, 53-56. Retrieved from: http://w3.umons.ac.be/staff/Delvaux.Veronique/vd/ICSLP2002.pdf

Flege, J. E. 1995. Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research*. (pp. 233–277). Timonium, MD: York Press.

Inceoglu, S. (2016). Effects of perceptual training on second language vowel perception and production. *Applied Psycholinguistics, 37*, 1175-1199.

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. Retrieved from https://search.proquest.com/docview/85513131?accountid=12846

Levy, E.S., & Law, F.F. (2010). Production of French vowels by American-English learners of French: Language experience, consonantal context, and the perception-production relationship. *Journal of the Acoustical Society of America, 128*(3), 1290-1305.

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.

Munro, M. J. & Derwing, T. M. (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech, 38*(3), 289-306.

Nishi, K., & Kewley-Port, D. (2008). Nonnative speech perception training using vowel subsets: Effects of vowel subsets and order of training. *Journal of Speech, Language, and Hearing Research, 51*, 1480-1493.

Rochet, B. (1995). Perception and production of second-language speech sounds by adults. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research* (pp. 379-410). Timonium, MD: York Press.

Saito, K. (2007). The influence of explicit phonetic instruction on pronunciation teaching in EFL settings: The case of English vowels and Japanese learners of English. *The Linguistics Journal, 2*(3), 16-40. Retrieved from https://search.proquest.com/docview/85660759?accountid=12846

Sheldon, A., & Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede perception. *Applied Psycholinguistics, 3*(3), 243-261.

Simon, E., Chambless, D., & Ubirata Kickhoefel, A. (2010). Understanding the role of orthography in the acquisition of a non-native vowel contrast. *Language Sciences, 32*(3), 380-394.

Thomson, R. I. (2012). Improving L2 listeners' perception of English vowels: A computer mediated approach. *Language Learning, 62*(4), 1231-1258.

Vandergrift, L. (2002). 'It was nice to see that our predictions were right': Developing metacognition in L2 listening comprehension. *The Canadian Modern Language Review/La Revue canadienne des langues vivantes, 58*(4), 555-575.

Vandergrift, L., & Tafaghodtari, M. H. (2010). Teaching L2 learners how to listen does make a difference: An empirical study. *Language Learning, 60*(2), 470-497.

Zampini, M. L. (2008). L2 speech production research: Findings, issues, and advances. In M.L. Zampini & J.G. Hansen Edwards (Eds.), *Phonology and Second Language Acquisition* (pp. 219-249). Philadelphia: John Benjamins Publishing Company.

Bliss, H., Cheng, L., Schellenburg, M., Lam, Z., Pai, R., & Gick, B. (2017) Ultrasound technology and its role in Cantonese pronunciation teaching and learning. 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. 33-46). Ames, IA: Iowa State University.

ULTRASOUND TECHNOLOGY AND ITS ROLE IN CANTONESE PRONUNCIATION TEACHING AND LEARNING

Heather Bliss, University of Victoria
Lauretta Cheng, University of British Columbia
Murray Schellenburg, University of British Columbia
Zoe Lam, University of British Columbia
Raymond Pai, University of British Columbia
Bryan Gick, University of British Columbia

Over a decade of research has shown how using ultrasound imaging for biovisual feedback of tongue movement can help improve language learners' pronunciation. However, ultrasound can be challenging to implement in a classroom context, as it is typically best with small groups and requires specialized training. One possible solution is the use of ultrasound overlay videos, in which mid-sagittal ultrasound videos of tongue movements in speech are overlaid on videos of an external view of a speaker's head to create videos in which the movements of the face and the tongue are viewed simultaneously. In this paper, we report on a study investigating the use of ultrasound overlay videos as a pronunciation learning tool in Cantonese language classes. Using a blended learning paradigm, half of the students interacted with the videos online to learn about two challenging sets of sounds: unreleased final stops and low central vowels, while half interacted with audio-only media under otherwise identical conditions. Results show that students who received the ultrasound-based training performed better in perception and production tasks.

INTRODUCTION

There is a growing body of evidence to support the use of ultrasound imaging technology in L2 learning (Cleland, Scobbie, Nakai, & Wrench, 2015; Gick, Bernhardt, Bacsfalvi, & Wilson, 2008; Noguchi et al., 2015; Ouni, 2014; Pillot-Loiseau, Kamiyama, & Kocjančič Antolík, 2015; Tateishi & Winters, 2013; Tsui, 2012; Wilson, 2014; Wilson & Gick, 2006; Wu, Gendrot, Hallé, & Adda-Decker, 2015). Ultrasound is useful in pronunciation teaching and learning, as it displays internal articulatory processes and facilitates the explanation and understanding of how to pronounce certain L2 sounds. However, there are various challenges associated with incorporating ultrasound into L2 learning programs. Specifically, given the current technology, ultrasound lends itself best to small groups or even individual learners. Moreover, raw, unedited ultrasound images may be difficult for untrained learners (and/or instructors) to interpret. (For a review of the body of literature on ultrasound-based pronunciation training, see Bliss, Abel, & Gick, 2017).

To address these challenges, a team of researchers at the University of British Columbia developed a technique for creating ultrasound overlay videos, in which mid-sagittal ultrasound videos of tongue movements in speech were overlaid on videos showing an external view of a speaker's head to create videos in which the movements of the face and the tongue are viewed simultaneously. The videos can be viewed at http://enunciate.arts.ubc.ca. There are 91 videos in the video library, corresponding to most of the sounds in the International Phonetic Alphabet. Designed for application in blended or online learning paradigms, the videos have been incorporated into introductory linguistics courses as well as some language courses at the University of British Columbia.

Although ultrasound is well-established as an effective interactive biofeedback tool in L2 pronunciation learning, there is yet little research on the effectiveness of ultrasound overlay videos. The types of feedback provided by real-time imaging of one's own productions versus those provided by overlay videos may be substantially different, and it is an empirical question whether ultrasound overlay videos can also benefit learners. Survey data suggests that learners enjoy the videos and feel they help for learning new sounds (Tsuda, Kim, Gick, Kazama, Yamane & Burton, 2015; Yamane et al., 2015). Also, a controlled experiment comparing the impact on linguistics student learning after exposure to different types of online tutorial materials found that student performance was better with tutorials that included ultrasound overlay videos than tutorials that included text materials (Abel et al., 2016). However, this current study is the first to systematically evaluate the effectiveness of ultrasound overlay videos as an L2 pronunciation learning tool.

While the video library of individual sounds is useful as a general resource, there is also a need for customized videos, in some cases targeting specific phonological contrasts in a given language, and in other cases to present a familiar face in a particular linguistic community. In this paper, we report on a pilot project to develop and evaluate ultrasound overlay videos focused on Cantonese words and sounds.

The reasons for focusing on Cantonese is twofold. First, the Cantonese language program at the University of British Columbia is new and in need of teaching materials. Launched in 2015, the program is the first of its kind in Canada; while Mandarin classes are widely offered at Canadian universities, the University of British Columbia's program is the first and only postsecondary for-credit Cantonese language program in Canada. Second, the phonetic and phonological properties of Cantonese pose specific pronunciation challenges for learners, some of which are likely to benefit from ultrasound-based instruction. Specifically, with regard to consonants, unaspirated stops /p/, /t/, /k/ are unreleased in coda position and as such, there is no release burst with these consonants to provide a perceptual cue to the place of articulation (Cheung, 1986). The formant frequencies of the preceding vowel are said to provide an acoustic cue to the identity of these consonants (e.g., Ciocca, Wong, & So, 1994; Khouw & Ciocca, 2006), but despite these acoustic cues, the unreleased stops can be difficult to distinguish, all "tending to sound like a glottal stop to an English speaker" (Matthews & Yip, 2011). With regard to vowels, Cantonese has two central low vowel phonemes [a:] and [ɐ], and given their close proximity, these two vowels are difficult for learners to distinguish.

Our research question is as follows: *Does interacting with ultrasound-enhanced videos improve beginner Cantonese learners' ability to differentiate between challenging Cantonese sounds in their perception and production?* We hypothesize that students who interact with ultrasound overlay videos will perform better in perception and production tasks that differentiate unreleased obstruents [p̚], [t̚], [k̚] and central low vowels [ɐ] and [a:] than students who interact with audio samples alone. In what follows, we describe an experiment that tested this hypothesis, beginning with a discussion of the methodology, followed by results, discussion, and concluding remarks.

METHODOLOGY

In this study, we developed ultrasound overlay videos for minimal sets of words that isolated the two contrasts, namely unreleased stops and central low vowels. We then conducted a comparative study in which half the learners were given access to the ultrasound overlay videos, and half were given access to the corresponding audio files. Learners were subsequently tested on their ability to differentiate the sounds in perception and production.

The participants were 13 undergraduate students enrolled in CNTO 301 (Basic Cantonese I) at the University of British Columbia. This course is an elementary level Cantonese language course designed for non-heritage learners with no prior exposure to or background in Cantonese. The focus of the course is on training for basic oral skills in the language. The 13 students were randomly assigned to two groups; an experimental group (n=7) received the ultrasound overlay videos and a control group (n=6) received audio files. The test items consisted of minimal sets contrasting the vowels and consonants, as shown in Tables 1 and 2.

Table 1

Test items: Vowels

Word	Transcription	English Translation
新	s**ɐ**n1	'new'
山	s**a:**n1	'mountain'
諗	l**ɐ**m2	'think'
攬	l**a:**m2	'hug

Table 2

Test items: Consonants

Word	Transcription	English Translation
濕	sɐ**p**1	'wet'
失	sɐ**t**1	'lose'
塞	sɐ**k**1	'(traffic) jam'
插	ts^{h}a:**p**3	'insert'
擦	ts^{h}a:**t**3	'erase'
拆	ts^{h}a:**k**3	'pull down / disassemble'

The words in Tables 1 and 2 were used to create stimuli consisting of ten videos and ten corresponding audio clips. The videos were produced by filming the side profile of a native Cantonese speaker (using a Panasonic Camcorder) while at the same time recording the ultrasound information from her tongue (using a CHISON portable ultrasound machine, recorded into an Mac Pro laptop using iMovie). The speaker was 34 years old at the time of the study, and had spent 5 years in Canada. She was born in Hong Kong and speaks the Hong Kong variety of Cantonese, along with English and Mandarin as additional languages. The speaker sat with the ultrasound probe positioned under the chin and read each target word three times. Two takes were filmed, and one example of each word – deemed by a native Cantonese speaker to be the clearest production – was selected to be developed into the ultrasound overlay video. Once recorded, the videos were time-aligned and the ultrasound videos trimmed, shaded, and overlaid on the videos of the face using Adobe Premiere (see Abel et al., 2015; Yamane et al., 2015 for a detailed description of the video production methods). The audio signal from each video was extracted and saved as a wav file. In each media file (video or audio) a single token was repeated three times, and the second repetition was manipulated to be 1.5 times slower.

The learners interacted with the stimuli through two near-identical tutorial websites created using WordPress. The experimental group was given access to a site that linked to the videos, and the control group was given access to the same site, but with links to the audio clips. On both sites, the media files were displayed alongside pictures that gave a graphical depiction of each word, as well as the word itself in the Cantonese orthography. The rationale for this type of display was based on the fact that students were not trained in how to read the Cantonese orthography, and they likely would not know the meaning of the words, but the images could make it clear that the media files corresponded to real Cantonese words and represented real contrasts. An example (from the experimental group's site) is given in Figure 1. (The top diagram in Figure 1 corresponds to the Cantonese word for 'to insert,' and the bottom diagram corresponds to the Cantonese word for 'to erase.')

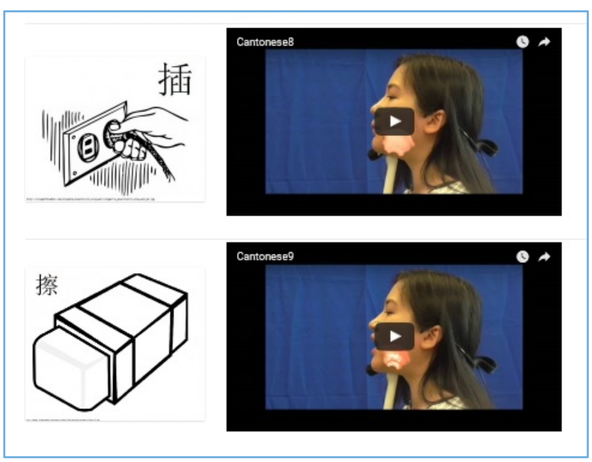


Figure 1. *Screenshot of example stimuli*

Each site was comprised of five pages, including a homepage plus one page for each minimal set. The pages were accessed via hyperlinks at the top of each page labelled “Set 1,” “Set 2,” etc. At the top of each page, a message directed students to focus on the phonetic contrast in question, as follows: “*Watch each video to hear the word corresponding to the picture. The word will be played three times, where the second repetition is slowed down. Pay special attention to the vowel sounds in the middle of the words.”* Access to the tutorial websites was through the University’s Online Learning Management System.

Following a week-long period of interacting with the websites at their leisure, learners were evaluated on their perception and production of the Cantonese sounds in question. Regarding the perception evaluation, this was administered as a post-test only through an online quiz with ten multiple choice questions that had accompanying audio files, recorded by a second native Cantonese speaker. This speaker was 21 years old at the time of the study, and had lived in Canada for 2.5 years. She was born in Hong Kong and speaks the Hong Kong variety of Cantonese, along with English, Mandarin, Italian, and Japanese. Half of the questions in the perception quiz were forced choice questions, with the following phrasing: “*Listen to this word. Does it correspond to the picture on the left or on the right*?” The other half of the questions were AXB questions; two of these focused on vowels, and the remaining three on consonants. The phrasing for the latter was as follows: “*Listen to these 3 words. If the consonant at the end of the first word is the same as the one at the end of the middle word, choose 1. If the consonant at the end of the last word is the same as the one at the end of the middle word, choose 3*.”

Regarding the production evaluation, both before and after the students accessed the tutorial websites, we recorded their productions of the numerals one through ten in both random and sequential orders. The recordings were carried out in a quiet classroom using a USB microphone and Audacity 2.1.2, and the recording conditions were identical in both the pre- and post-recording sessions. The numerals were chosen because the students were familiar with these words and therefore would not need to repeat after their instructor or another native speaker, and the numerals contain the relevant phonological contrasts, as illustrated in Table 3.

Table 3

Numerals illustrating relevant contrasts

Vowels	Consonants
3 三 [s**aː**m1]; 8 八 [b**aː**t3]	10 十 [sɐ**p**6]
10 十 [s**ɐ**p6]; 7 七 [tsʰ**ɐ**t1]	7 七 [tsʰɐ**t**1]
	6 六 [lo**k**6]

With respect to data analysis, the perception data was analysed according to the percentage of correct responses by question type across conditions. For the production data, we performed both acoustic and rater analyses. Acoustic analysis was carried out using Praat (Boersma & Weenick, 2015), and analysis of the vowels entailed collecting F1 values of the vowels at midpoint and comparing across vowels and conditions. Analysis of the consonants built upon a method used by Ciocca et al., 1994 and Khouw & Ciocca, 2006, who measured F1, F2, and F3 values at three durational points for Cantonese final consonants. We additionally calculated the difference between F2 and F3 values from the midpoint to the endpoint of vowel duration, and compared across pre- and post-recordings for each subject and across conditions. The rater analysis entailed having four native Cantonese speakers rate the nativeness of the vowels and consonants separately on scale of one to five. The rating task involved rating 540 tokens, which included four different productions of the same word by the same participant. Each sound file was played and rated once. We compared mean rater scores for vowels and consonants using two-way random intra-class correlations as a measure of interrater reliability.

RESULTS

Due to a small sample size, statistical analysis was not possible for most of the results reported in this section. Instead, we comment on the observed trends and whether these trends are in the predicted direction. To begin, the results of the perception quiz are presented in Table 4.

Table 4

Perception Results

Group	Mean Correct Responses (AXB)	Mean Correct Responses (Forced Choice)
Experimental	88% (n=5)	70% (n=6)
Control	77% (n=6)	63% (n=6)

Statistical analysis of the perception results is not reliable due to the small number of participants. We omitted one outlier whose score was significantly lower than all other participants, as well as a participant from the experimental group who did not answer the AXB questions. The results of the acoustic analysis of the vowels are presented in Figure 2 (in which “aa” refers to [a:] and “a” refers to [ɐ]).

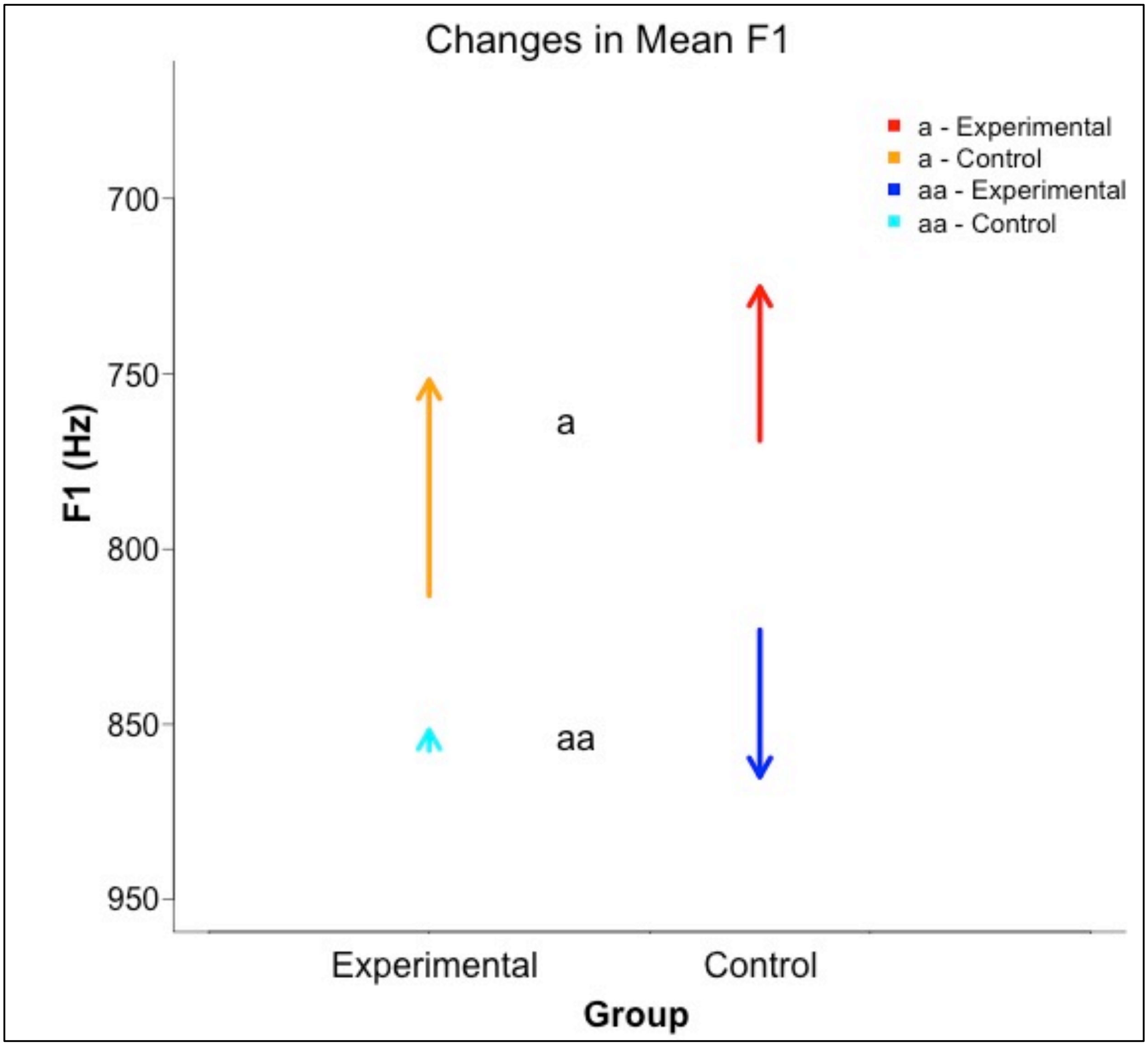


Figure 2. Acoustic measures of the vowels

The arrows in Figure 2 show the mean direction of change in F1 values for each vowel from pre- to post-recording. For native speakers (including the two used as model speakers in our study), [ɐ] (represented by ‘a’) is produced higher in the mouth than [a:] (represented by ‘aa’), and correspondingly has higher F1 values, but learners often have trouble with this distinction. As learners improve their pronunciation, we expect them to differentiate between the vowels, with a higher F1 value for [ɐ] than [a:]. In Figure 2 we see that, while both control and experimental groups seem to have a larger difference between the F1 values of the two vowels after training, the experimental group show trends towards a larger increased differentiation of the two low vowels than the control. In other words, comparing across time points, the differences between the height of the tongue for [ɐ] and [a:] is more pronounced in the post-recording sessions regardless of condition. Comparing across conditions, the experimental group appears to have improved more. Formant values were not normalized because we were focusing on changes within subjects, not raw formant values. The result of the rater analysis for the vowels are presented in Table 5; the rating scale was 1-5, with 5 being most nativelike.

Table 5

Rater Analysis (Vowels)

Group	pre-/post-	Mean Rating (all)	Mean Rating [ɐ]	Mean Rating [a:]
Control	pre-	3.565	3.563	3.568
	post-	3.456	3.573	3.339
Experimental	pre-	3.784	3.755	3.813
	post-	3.456	3.568	4.078
Native	--	4.555	4.650	4.517

Two-way random intra-class correlations were run as a measure of interrater reliability, indicating fair agreement amongst the four raters [ICC(2, 4) = 0.455]. Table 5 shows the mean ratings between conditions and recordings for each of the vowels and for all vowels together. A mixed ANOVA was conducted on the ratings between the variables of condition and recording, finding significant variation among conditions ($F = 15.22$, $p < 0.01$). A post hoc Tukey test showed that there were significant differences between ratings of native speakers and learners ($p < 0.01$), but not between conditions in the Cantonese learners. In other words, there were significant differences between ratings of native speakers and learners, but not between conditions in the Cantonese learners. Nevertheless, there appears to be a slight trend towards an increase in nativeness rating for the experimental group's production of [a:], which matches the trend of improvement from acoustic data.

Turning to the consonants, we found no significant differences in the F2 or F3 values between the different consonants, even for the native Cantonese model speakers. Similarly, there were no significant changes in the ratings for either group over time, as shown in Table 6. As with the vowels, measures of interrater reliability indicate that there was fair agreement amongst the four raters [ICC(2, 4) = 0.593]. While there were no trends between conditions, the data did show a pattern of nativeness ratings across the three stops whereby [p˺] was consistently rated highest and [k˺] lowest.

Table 6

Rater Analysis (Consonants)

		Mean Rating			
Group	pre-/post-	(all)	[p˺]	[t˺]	[k˺]
Control	pre-	3.607	4.104	3.487	3.596
	post-	3.662	3.729	3.544	3.281
Experimental	pre-	3.848	3.927	3.755	3.573
	post-	3.960	3.927	3.800	3.604
Native	--	4.450	4.750	4.156	4.594

Discussion

Recall our prediction that students who interact with the ultrasound overlay videos would perform better in production and perception of the unreleased stops and low central vowels than students who interact with audio files. While the small sample size in this study limits our ability to draw any firm conclusions, in a broad sense this prediction is borne out. The perception data trended in the right direction, with the experimental group showing more improvements than the control group, and the production data showing more mixed results, but with trends in the right direction (for the vowels particularly).

Regarding the perception data, we observed a smaller difference between the experimental and control groups with the forced choice questions, and we speculate that this may be because these questions did not test pronunciation so much as vocabulary, relying on students' ability to recall the training words. The overall weak effect in the perception data could be due to problems with both quiz design and delivery. With respect to quiz design, instead of forced choice questions, we speculate same/different types of questions ("*Are these two sounds the same or different*?") may have been more effective. With respect to quiz delivery, a technical glitch required that students type their responses (e.g., they had to type "left" or "right" in response to the forced choice questions and "1" or "3" in response to the AXB question), rather than simply by clicking a button to select from a set of options, and this may have impacted the results.

As for the acoustic results, as evidenced in Figure 2, the vowels trended in the right direction, with a clearer distinction in the experimental group between the F1 values of [a:] versus [ɐ]. The consonants, on the other hand, showed no acoustic differences between [p˺], [t˺], and [k˺], even for the model speaker. One possible explanation for this lack of a distinction is an ongoing sound change in Cantonese whereby younger speakers are fluctuating between word-final [t˺] and [k˺], using the two in free variation (Law, Fung, & Bauer, 2013; To, McLeod, & Cheung, 2015). It is possible that the model speaker may have not differentiated these sounds as clearly, which would have been reflected in stimuli the students received, a hypothesis that is supported by the observation that the F2 and F3 values of the consonants were not significantly different for any of the speakers. Alternatively, it could be that glottalization obscured the contrast. Particularly with younger speakers, Cantonese final unreleased stops can often have a glottal closure component, which means that the vocal folds close to stop airflow momentarily (Yip, 2015). The [p˺], [t˺], and [k˺] stops are often produced with glottal closure, before or during tongue movement, and sometimes even instead of any tongue articulation. Such a glottal closure would at least partially obscure the acoustic cues measured. If the tongue articulation was also weaker because of the glottal closure, this could explain the weak effects in even our model speaker data, which in turn would affect the participants' performance. The consequence of this extra articulation may be that ultrasound information alone is insufficient for learning these sounds.

Regarding the results of the native speaker ratings, the lack of a clear result may be related to task effects. Raters reported that being asked to judge only the vowel or consonant when hearing the full word to be a difficult task. Consequently, it could be the case that the ratings were inadvertently influenced by overall nativeness judgments, and/or from the interference of tone. Moreover, validity and interrater reliability may have been improved by allowing raters to listen to each token multiple times.

Some other factors may have impacted the results that we found. A software glitch prevented us from tracking the amount of time each student spent on the training website, and as a result we are unaware of any correlations between time spent on the websites and production and perception performance. As a result, it is not possible to know whether the encouraging trends reported here are due to the differential training across the groups or to differences in time spent on the training task between the two groups. At the time of writing, we are conducting a follow-up study that has measures in place to ensure training time can be tracked. Furthermore, the students had taken a midterm in which they were tested on the numerals the day before the pre-recording session. The post-recording session was one week later. This timing may have influenced the students' production performance, as they would have been more well-studied at the time of the pre-recording session.

CONCLUSIONS AND FUTURE DIRECTIONS

In this study, we tested the use of ultrasound overlay videos as a pronunciation learning tool, in contrast with audio-only media. While the study focused on only a small set of sounds and had a small number of participants, the results are promising insofar as they trend towards providing empirical support for the use ultrasound overlay videos for learning certain Cantonese sounds that are reportedly challenging for L2 learners. This study serves a pilot, and at the time of writing we are in the process of conducting a follow-up study with Cantonese learners that addresses some of the problems identified of the current study, and includes a larger number of participants. We are also developing customized videos for other widely taught languages at the University of British Columbia, including Mandarin, Spanish, French, and German, and we are responding to interest from other linguistic communities and making customized videos for languages including SENĆOŦEN, Blackfoot, Secwepemc, and Halq'emeylem (Bliss, Burton, & Gick, 2016). As our video library grows, we continue to develop plans to integrate the videos into classrooms and other language learning environments, and to evaluate these videos' impact on pronunciation learning.

ACKNOWLEDGEMENTS

Sincere thanks to Strang Burton, Johnathan Chan, Deena Hymowitz, Matthew Law, Yik Tung Wong for their assistance with this project, to other members of the Interdisciplinary Speech Research Laboratory and the audience at PSLLT 8 for insightful comments and feedback, and to the students of CNTO 301 (Summer 2016) for their willing participation. This work has been supported by an NIH Grant DC-02717 to Haskins Laboratories, the UBC Teaching Learning Enhancement Fund, a UBC Arts Undergraduate Research Award, and a Robert and Kazuko Barker Award.

ABOUT THE AUTHORS

Heather Bliss is a Banting Postdoctoral Fellow at the University of Victoria. Having worked with Indigenous communities in Western Canada over the past fifteen years on language documentation projects, she has a background and interest in endangered language revitalization and community-based strategies for heritage language learning. Recently, she has been exploring the use of ultrasound technology as a tool for pronunciation teaching and learning, as well as language documentation, in Indigenous communities.

Department of Linguistics, University of Victoria
Clearihue Building D353
Victoria, BC V8W4P6
hbliss@uvic.ca

Lauretta Cheng is an undergraduate research assistant in the Department of Linguistics at the University of British Columbia. Her research interests are on phonetics in bilingual and heritage speakers, second language teaching and acquisition, as well as Cantonese linguistics. She has worked on projects involving L2 pronunciation teaching, bilingual articulatory compensation, and phonetic variation and change in heritage Cantonese.

Department of Linguistics, University of British Columbia
Totem Field Studios, 2613 West Mall
Vancouver, BC V6T1Z4
lauretta.cheng@alumni.ubc.ca

Murray Schellenberg is a post-doctoral researcher and manager of the Interdisciplinary Speech Research Laboratory (ISRL) at the University of British Columbia. His main research interests involve the interaction of language and music. He is particularly interested in issues related to the influence of language on the creation and/or performance of music. Topics of study include the realization of lexical tone in singing in tone languages (particularly in Cantonese and Mandarin); the carry-over of spoken accent to singing performance and influences of phonological structure on text setting.

Department of Linguistics, University of British Columbia
Totem Field Studios, 2613 West Mall
Vancouver, BC V6T1Z4
mhschellenberg@gmail.com

Zoe Lam is a PhD Candidate in Linguistics from the University of British Columbia. Her research interests include the perception of lexical tones, heritage linguistics, phonetics-phonology interface, and the grammar of Cantonese discourse particles. Currently her dissertation project focuses on the perception of Cantonese speech sounds by English-dominant heritage speakers in Canada who grew up in a Cantonese household. Zoe is also a social activist for linguistic diversity. She is an active member of the Chinatown community in Vancouver and is also the instructor of a Cantonese Saturday School.

Totem Field Studios, 2613 West Mall
Vancouver, BC V6T1Z4
zoelam@alumni.ubc.ca

Raymond Pai is lecturer and the director of the Cantonese program at the University of British Columbia. He is recently named Lecturer of the Year in the Department of Asian Studies at UBC. He taught Chinese at the Defense Language Institute Foreign Language Center and the Middlebury Institute of International Studies in Monterey, California for nearly a decade. He is also the Cantonese examiner for the University of Arizona. His research focus is the language policies in Hong Kong and Cantonese language pedagogy. His other academic and research interests include language learning motivation, language attitudes, language technology and testing, and psycholinguistics.

Department of Asian Studies, University of British Columbia
607 – 1871 West Mall
Vancouver, BC V6T1Z2
kitpai@mail.ubc.ca

Bryan Gick is Professor and Head of the Department of Linguistics at University of British Columbia and a Senior Researcher at Haskins Laboratories. His research interests have focused mainly on understanding the physical mechanisms of speech production and their interactions with perception, phonology and phonetics in the speech of normal and disordered adults and children. He has developed techniques for applying ultrasound imaging technology to speech research, protocols for studying tactile perception of speech, methods for observing low-level planning in speech motor behaviour, and computational tools for biomechanical simulation.

Department of Linguistics, University of British Columbia
Totem Field Studios, 2613 West Mall
Vancouver, BC V6T1Z4
gick@mail.ubc.ca

REFERENCES

Abel, J., Allen, B., Burton, S., Kazama, M., Noguchi, N., Tsuda, A., Yamane, N., & Gick, B. (2015). Ultrasound-enhanced multimodal approaches to pronunciation teaching and learning. *Canadian Acoustics, 43*(3), 124-25.

Abel, J., Bliss, H., Noguchi, M., Schellenberg, M., Yamane, N., & Gick, B. *Instructional reinforcements in phonetics pedagogy: A comparative study*. (2016). Unpublished manuscript (under review), the University of British Columbia.

Bliss, H., Abel, J., & Gick, B. (2017). Computer-assisted visual articulation feedback in L2 pronunciation instruction: A review. Unpublished manuscript (under review), the University of British Columbia.

Bliss, H., Burton, S., & Gick, B. (2016). Ultrasound overlay videos and their application in Indigenous language learning and revitalization. *Canadian Acoustics, 44*(3), 136-37.

Boersma, P. & Weenink, D. (2015). Praat: doing phonetics by computer. Version 5.4.08, Available from http://praat.org/

Cheung, K.H. (1986). *The phonology of present-day Cantonese*. PhD dissertation, University College London.

Ciocca, V., Wong, L., & So, L. K. H. (1994). An acoustic analysis of unreleased stop consonants in word final positions. *Proceedings of the International Conference on Spoken Language Processing, Yokohama, 21*, 1131–1134.

Cleland, J., Scobbie, J.M., Nakai, S., & Wrench, A. (2015). Helping children learn non-native articulations: the implications for ultrasound-based clinical intervention. Paper presented at the 2015 International Conference of Phonetic Sciences, Glasgow, Scotland. Retrieved on 12 August 2015 from http://www.icphs2015.info/pdfs/Papers/ICPHS0698.pdf.

Gick, B., Bernhardt, B., Bacsfalvi, P., & Wilson, I. (2008). Ultrasound imaging applications in second language acquisition. In Hansen Edwards, J. G., & Zampini, M. L. (Eds.). *Phonology and second language acquisition* (pp. 309-322). Amsterdam: John Benjamins.

Khouw, E., & Ciocca, V. (2006). An acoustic and perceptual study of final stops produced by profoundly hearing impaired adolescents. *Journal of Speech, Language, and Hearing Research*, *49*(1), 172-185.

Law, S.P., Fung, R.S.Y., & Bauer, R.S. (2013). Perception and production of Cantonese consonant endings. *Asia Pacific Journal of Speech, Language and Hearing, 6*(3), 179-195.

Matthews, S., & Yip, V. (2011). Unbalanced bilingual acquisition as a mechanism of grammatical change. *Bilingualism: Language and Cognition, 14*(02), 159-161.

Noguchi, M., Yamane, N., Tsuda, A., Kazama, M., Kim, B., & Gick, B. (2015). Towards protocols for L2 pronunciation training using ultrasound imaging. Poster presented at PSLLT 2015. Dallas, Texas.

Ouni, S. (2014). Tongue control and its implication in pronunciation training. *Computer Assisted Language Learning, 27*(5), 439-453.

Pillot-Loiseau, C., Kamiyama, T., & Kocjančič Antolík, T. (2015). French /y/-/u/ contrast in Japanese learners with/without ultrasound feedback: vowels, non-words and words. Paper presented at the 2015 International Conference of Phonetic Sciences, Glasgow, Scotland. Retrieved on 12 August 2015 from http://www.icphs2015.info/pdfs/Papers/ICPHS0485.pdf.

Tateishi, M., & Winters, S. (2013). Does ultrasound training lead to improved perception of a non-native sound contrast? Evidence from Japanese learners of English. Paper presented at the 2013 meeting of the Canadian Linguistic Association, Victoria, BC, Canada. Retrieved on 12 August 2015 from http://homes.chass.utoronto.ca/~cla-acl/actes2013/Tateishi_and_Winters-2013.pdf.

To, C.K.S., McLeod, S., & Cheung, P.S.P. (2015) Phonetic variations and sound changes in Hong Kong Cantonese: Diachronic review, synchronic study and implications for speech sound assessment. *Clinical Linguistics & Phonetic*s, *29*(5), 333-353.

Tsuda, A., Kim, B., Gick, B, Kazama, M., Yamane, N, & Burton, S. (2015). *Ultrasound-integrated pronunciation tutorials.* Roundtable discussion presented at the Society for Teaching and Learning in Higher Education, Vancouver, BC.

Tsui, H.M. (2012). *Ultrasound speech training for Japanese adults learning English as a second language*. Unpublished MSc thesis, University of British Columbia.

Wilson, I. (2014). Using ultrasound for teaching and researching articulation. *Acoustical Science and Technology*, *35*(6), 285-289.

Wilson, I., & Gick, B. (2006). Ultrasound technology and second language acquisition research. In Grantham O'Brien, M., Shea, C., & Archibald, J. (Eds.). *Proceedings of the 8th Generative Approaches to Second Language Acquisition Conference (GASLA 2006)* (pp. 148-152). Somerville, MA: Cascadilla Proceedings Project.

Wu, Y., Gendrot, C., Hallé, P., & Adda-Decker, M. (2015). On improving the pronunciation of French /r/ in Chinese learners by using real-time ultrasound visualization. Paper presented at the 2015 International Conference of Phonetic Sciences, Glasgow, Scotland. Retrieved on 12 August 2015 from http://www.icphs2015.info/pdfs/Papers/ICPHS0786.pdf.

Yamane, N., Abel, J., Allen, B., Burton, S., Kazama, M., Noguchi, M., Tsuda, A., & Gick, B. (2015). Ultrasound-integrated pronunciation teaching and learning. Paper presented at Ultrafest VII, University of Hong Kong.

Yip, J. (2015). Gestural reduction of Hong Kong Cantonese syllable-final oral stops. Proceedings of the International Conference of Phonetic Sciences, Glasgow, Scotland. Retrieved on 11 August 2016 from http://www.icphs2015.info/pdfs/Papers/ICPHS0912.pdf.

Dickerson, W. (2017). The baby in the rhythmic bathwater. 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. 47-57). Ames, IA: Iowa State University.

THE BABY IN THE RHYTHMIC BATHWATER

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

TESL/TEFL professionals are generally aware that the version of stress-timed rhythm commonly taught has been thoroughly discredited by linguistic research: Native English speakers do not accent every content word nor do they deliver accents at regular intervals. To do so is to speak an unknown variety of English that challenges listeners' ability to understand. Less well known is that another model of English rhythm has existed for as long as the term *stress-timed rhythm* has. This model comes from phoneticians' fieldwork on spontaneous speech and describes a variety of English in actual use.

This paper addresses instructors interested in teaching a more authentic English. Recognizing that stress-timed rhythm and pedagogical practices promoting it have some merit, it seeks to (a) separate the valuable parts of the stress-timed model from the worthless parts—the proverbial "baby" from the "bathwater;" (b) expand on the valuable parts to describe the alternative model more fully; and (c) identify ways to integrate the alternative model into pronunciation teaching even when teachers are using textbooks built around stress-timed rhythm.

INTRODUCTION

Most British, American, and Australian ESL/EFL pronunciation textbooks claim to teach a certain regional, social, or functional variety of English (Hewings, 2007; Grant, 2016; Zawadzki, 1994). Their target is nonetheless implicitly, if not explicitly, spontaneous speech. While attention to prosody and connected speech has moved textbooks closer to ordinary talk, they are unlikely to reach this target by continuing to teach the traditional model of rhythm—stress timing, for which phoneticians have found no empirical support (Arvaniti, 2012, pp. 351-53). Phrases spoken with this non-English rhythm undermine intelligibility by making it harder for listeners to understand and remember meaning (Hahn, 1999; Dickerson, 2016). To prepare learners to use everyday English, we need to teach them a more accurate variety of its foundational rhythm.

In Dickerson (2015), I describe the dominant rhythm of unrehearsed speech, namely, the *two-peak profile,* and trace the 70-year history of British and American research supporting it.[1] I also show how to make this model accessible to linguistically unsophisticated learners of English. This paper addresses a major impediment to progress in English rhythm pedagogy: If stress timing does not accurately represent English, and the two-peak profile does, how can ESL/EFL instructors teach the rhythm of English using a textbook that promotes stress timing, as most do?

[1] Kenneth Pike is the earliest phonetician I have found who identified the two-peak profile (Fries 1945, p. 64). (References in Fries are to Pike's work.). Other researchers who have described the two-peak profile as characteristic of spontaneous speech are Bolinger (1961), Brazil, et al. (1980), Cauldwell (2002), and Wells (2006).

AGAINST THE GRAIN

Whenever an improvement in ESL/EFL pedagogy comes along—a model or approach—the history of our field identifies our inclination. At some point we adopt the new and reject the old (Murphy & Baker, 2015). The two-peak profile is one such improvement. While not actually new, its implementation for ESL/EFL instruction is. Given our tendency to discard the old—in this case stress-timed rhythm, this paper cautions against tossing out the BABY (enduring concepts) with the bathwater (unsupported concepts)—lest we reject things of value to the two-peak profile when rejecting the repudiated model. Finding the real worth in stress timing is the place to start.

Over a period of three academic years, I transitioned my pronunciation course from an emphasis on stress-timed rhythm to the more faithful two-peak model. In the process, Laura Hahn and I rewrote *Speechcraft*, our pronunciation text (Dickerson & Hahn, forthcoming). What follows summarizes the issues we addressed during this transition.

TWO MODELS

The constructed examples in Figure 1 illustrate how the two models of English rhythm would mark accents. The filled-in bullet (●) we call the **primary peak**. Bolinger (1961) calls it an *accent,* Halliday (1967), the *tonic,* and Wells (2006), the *nucleus*. In pedagogical texts it goes by *prominence* (Celce-Murcia, et al., 2010), *focus word* (Gilbert, 2012; Grant, 2017), and *primary stress* (Hahn & Dickerson, 1999). The hollow bullets (○) to its left, if any, are accents of a lesser degree. In the two-peak profile, the single hollow bullet is the **anchor peak,** also known as the *onset* (Wells, 2006). These accents combine rhythm and intonation and return to an approach to rhythm advocated by Pike (1945), where accented syllables begin intonation contours in speech. All other words and syllables in a phrase compose what we call **valleys.**

Stress-Timed Rhythm		**Two-Peak Profile**
I live seven blocks from campus.	1	I live seven blocks from campus.
His popular seminar is rarely canceled.	2	His popular seminar is rarely canceled.
Which library has the best facilities?	3	Which library has the best facilities?
She’s an engaging public speaker.	4	She’s an engaging public speaker.
You could have told me about him.	5	You could have told me about him.

Figure 1. Accents according to TESL’s stress-timed rhythm and the two-peak profile.

In the examples above, the same rule places the primary peak in both models. Different rules assign lesser accents. According to the model of stress timing commonly found in ESL/EFL materials (left column above), a lesser accent goes on every lexical word—noun, adjective, verb, and adverb—and certain categories of grammatical words, namely, information-question words, demonstrative subject pronouns, and indefinite subject pronouns, collectively called *content* *words*. All remaining grammatical words are *function words* (Pike, 1945).[2]

We (Dickerson, 2004; Hahn & Dickerson, 1999) find it helpful to relabel these groupings. We use the term *content word* for the four categories of lexical words. The grammatical categories that Pike includes with content words, we call *loud function words.* To this group we add negative words. The words that Pike calls *function words,* we label as *soft function words.* See Table 1.

Table 1

Divisions of the English lexicon in Dickerson, (2004); Hahn & Dickerson, (1999)

Content Words	**Loud Function Words**	**Soft Function Words**
Nouns	Negative words and contractions	Articles, pronouns, prepositions, auxiliaries, modals, conjunctions, etc.
Adjectives	Information question words	
Verbs	Demonstrative subject pronouns	
Adverbs	Indefinite subject pronouns	

The two-peak profile has at most only one lesser accent per phrase—the anchor. It is assigned from the beginning of the phrase after the primary peak is in place. The anchor prefers words with the highest semantic salience among content words and loud function words, specifically, number words (#) (example 1 in Figure 1 above), **n**ouns (example 2), **a**dverbs, **i**mperative verbs. All **l**oud function words are also anchor preferences (example 3). A memory aid for these preferences is the acronym **#NAIL** (Dickerson, 2015)**.** The anchor settles on whichever preference comes first in a phrase. At times you must pass over a word or two to find a preference. If no #NAIL word occurs left of the primary peak, the anchor will default to the first adjective or verb (**1st Aj/V**), if available (example 4). If not available, then there is no anchor (**Ø**) (example 5). The anchor-placement decision procedure is abbreviated as #NAIL, 1st Aj/V, Ø (null).[3] These are the ordered steps of the rule:

1. Numbers, nouns, adverbs, imperative verbs, loud function words (in any order)
2. 1st Adjective or Verb
3. Otherwise, no anchor

[2] Pike does not list negatives. Prator & Robinett make negatives explicitly part of content words in 1985, p. 31.

[3] Pike also found that native speakers regularly hop over non-imperative verbs and non-number adjectives: "In a short subject-verb-complement phrase, the verb is often left unstressed" (1945, p. 118): *They gave us fruit for our dessert* (Fries, 1943, p. 293). He also deaccents phrase-initial adjectives: *There's a small grocery store on State Street* (Fries, 1943, p. 202). Lacking a #NAIL word, the anchor defaults to the first adjective or verb, if available, e.g., *He's writing a large book* (Fries, 1943, p. 292). If unavailable, there is no anchor, e.g., *Does he want it?* (Fries, 1943, p. 222).

THE BABY IN THE BATHWATER: WHAT ENDURES

With both models and their rules for assigning lesser accents (○) before us, we can now identify the features and classroom activities that stress-timed rhythm got *right* because they apply equally to the two-peak profile. This reveals a partial picture of the BABY; the rest will appear when we discuss alternatives to the discarded bathwater.

First, both models claim that the neutral position of peaks and valleys is predictable. This reassuring claim says to learners: "There's a pattern to follow; you don't have to guess."

Second, as noted, both models select words for peaks and valleys based on their semantic salience, using Pike's *content word* / *function word* dichotomy as defined by parts of speech.

Third, in both models only one syllable is most-prominent (the primary peak) in a non-contrastive (neutral) phrase. It occurs on the same syllable of the same word in both models. This default placement is the final content word in the phrase.

Stress-Timed Rhythm	**Two-Peak Profile**
I live seven blocks from campus.	I live seven blocks from campus.

Fourth, both models are capable of signaling emphasis with the primary peak because it can override any other degree of stress. Here the primary peak moves to the word *best.*

Which library has the best facilities?[4]	Which library has the best facilities?

Fifth, in both models some neutral phrases have only one primary peak and no other accents. Such phrases have neither stress timing nor two peaks, despite the names of their models.

You could have told me about him.	You could have told me about him.

Sixth, instructors of both models help learners create their respective rhythms by contrasting peaks and valleys to make peaks stand out, and compressing valleys to reduce their size.

These six comparisons show that when we taught stress timing, some features also applied to the two-peak profile. These commonalities might suggest that stress timing *almost* got English rhythm right. That is not the case. There is much more to the BABY than these six points.

THE REST OF THE BABY: WHAT REPLACES THE BATHWATER

To talk about the differences between the two models, and what must be tossed out, we also need to identify what replaces each rejected notion and adds to the picture of the BABY.

Seventh, the two models use the word *rhythm* differently. For stress timing, rhythm is regular. Predictable beats make valleys the same size. We now know that the impression of regular timing is a perceptual illusion (Cauldwell, 2002). In the two-peak profile, rhythm is unrelated to time; its

[4] To emphasize a particular word, Pike says "stress that word and eliminate the normal stress in the sentence" (Fries, 1943, pp. 175-76).

valleys are different sizes. Rhythm is in the repeated alternation between peaks and valleys. This *non-temporal sense of rhythm* is part of the BABY; equally paced accents are bathwater.

Eighth, we do not put an accent on every content word in a phrase. Prator (1951, pp. 25-26) introduced this accent rule 65 years ago to help his ESL students predict accents, despite Pike's repeated insistence to the contrary.[5] The infrequent use of *peaks in a phrase—at most only two—* is a feature of the BABY. Accenting every content word is bathwater.

Ninth, the rationale for compressing valley words is different. Stress-timed rhythm claims that we compress to keep peaks coming at an even tempo (Fries, 1943, p. 200). Given that peaks do not occur at regular intervals, there must be another reason to compress valleys.

Wells (2006, p. 234) says that the words from the anchor peak through the primary peak form a semantic unit he call the *focus domain* or *foreground* of the speaker's message.[6]

She's an [engaging public speaker].

Solely producing peaks in a phrase does not ensure that listeners will hear the accented words as a unit. Olle Kjellin says that the two peaks bounding the unit must also be close enough together (1999, p. 12). To draw peaks together, speakers must use (a) a rapid delivery having the expected rhythm and (b) valley compression. *Squeezing valley syllables to improve the comprehensibility of a message* is a meaning-based reason to compress valleys and part of the BABY. Compression to maintain a regular tempo is bathwater.

Tenth, the consequence of not accenting every content word is that lexical words fill valleys generally making them longer (on the right below.) *Suppressing lexical stress* is a dominant feature of spontaneous spoken English and part of the BABY (Wells, 2006, p. 229).

I live seven blocks from campus.	1	I live seven blocks from campus.
His popular seminar is rarely canceled.	2	His popular seminar is rarely canceled.
Which library has the best facilities?	3	Which library has the best facilities?
She's an engaging public speaker.	4	She's an engaging public speaker.
You could have told me about him.	5	You could have told me about him.

Figure 2. Comparison of valleys in stress-timed and two-peak profile phrases.

Eleventh, while both models of rhythm can signal emphasis with the primary peak, only the two-peak profile can signal emphasis with the anchor peak. A lesser accent on every content word in

[5] In Fries (1943), Pike demonstrates that *no* class of words always carries a peak or is always in a valley. The prominence of a word depends on the meaning highlighted or backgrounded by the rhythm of a phrase.
[6] Brazil, et al. (1980, pp. 39-42) call this unit the *tonic segment.* It carries all the intonational meaning of the phrase.

stress-timed phrases precludes moving or adding an accent (left-hand example below). In the two-peak profile, moving or adding anchors for emphasis is not only possible; it is also common. On the right, the neutral anchor is given first.

Now imagine talking about a professor's popular and not-so-popular seminars. When you want to comment on his popular seminar, you move the anchor to *popular* to contrast it with the *unpopular* seminar, as in the second phrase on the right. The *flexibility to emphasize with the anchor* is another part of the BABY.

Stress-Timed Rhythm	**Two-Peak Profile**
His popular seminar is rarely canceled.	His popular seminar is rarely canceled.
	His popular seminar is rarely canceled.

The **twelfth** contrast concerns the length of spoken phrases. Phrase length is irrelevant to stress-timed rhythm; speakers accent each new content word, however long the phrase. By contrast, the two-peak profile is a feature of only one style of speech, namely, spontaneous speech. In this style, the limits of on-the-fly speech processing constrain phrases to about seven words.[7] *This word limit is integral to the two-peak profile,* and another critical part of the BABY.

This comparison of the BABY and the bathwater fills out the picture of the BABY and suggests how to teach the two-peak profile even when our textbook encourages stress timing. It also makes clear the underlying motivation for such instruction.

The prosodic features of spontaneous English speech—those we call the BABY—help listeners understand the speaker's meaning. Our pronunciation teaching should therefore *help learners help listeners*. We can help learners, first, by selecting appropriate rhythm content for their level of English. Second, we can provide appropriate practice materials and activities that promote their rhythm skills. We consider the second point first.

APPROPRIATE PRACTICE MATERIALS AND ACTIVITIES

A. Use short phrases and help learners create them. Short, simple phrases in the range of 1 to 7 words are a courteous length for listeners who are accustomed to hearing and processing language packaged this way. When learners speak spontaneously they naturally respect these limits. When they deliver memorized text or read from a script, they generally ignore them, challenging listeners' processing capabilities.

We can help learners become comfortable with the sound and feel of the two-peak profile if *all* practice phrases stay within the 1-7-word range. We should go farther than this. Many learners, especially burgeoning academics who often give presentations from a text, need explicit instructions on how to recast written phrases and add boundaries to simulate spontaneous speech. They also need periodic practice using these instructions.

B. Practice identifying parts of speech. Recognizing parts of speech is an essential skill, whether predicting the primary or the anchor peak. Most students can do this because of past instruction. Some, however, need extra help identifying the four types of lexical words and the

[7] Our study of 607 spontaneous phrases in two transcripts in Brazil, et al. (1980) shows that 95% and 94%, respectively, are within the 1-7-word range.

four types of loud function words. Teachers who offer such help can turn a highly-frustrated student into one who experiences a satisfying level of success using oral English.

C. Give prediction homework. In our experience, learners position peaks more accurately if they do out-of-class, pencil-and-paper exercises placing peaks by rule and then pronounce their predictions aloud. Such practice enables them to monitor and correct themselves better in class and out of class. In class, instead of correcting students' rhythm, we ask leading questions (queries) that guide them to self-correct. Out of class, these same queries and prediction rules are valuable internal resources for ongoing self-improvement. The positive long-term effects of using formal rules are confirmed in Sardegna's (2009) research.

D. Target rhythm basics in oral practice. Most learners do not intuitively use their voices to highlight and background vowels or to compress valley syllables. Therefore, in-class oral work should include regular practice contrasting the suprasegmentals of peak and valley vowels, and using assimilation, trimming, reduction, and linking to compress valley syllables.

Among rhythm basics, teachers should also be prepared to help learners sustain quick, quiet syllables across every word in a valley, whether lexical or grammatical. A tool often associated with stress timing can help. Instead of emphasizing valleys of uniform size, we can repurpose build-ups to increase learners' control of longer and longer valley strings. The following build-up focuses on the valley between peaks, allowing learners to extend its duration as syllables increase.

We often rehearse.

We often rehearse phrases.

We often rehearse our phrases.

We often rehearse our phrases aloud.

We find no value, however, in exercises that promote stress timing—rhymes, limericks, chants, and poems.

We are not suggesting that other exercises in stress-timed textbooks are also worthless. While students need work on the fundamentals in points A-D to become familiar with the new rhythm, their practice materials must incorporate these basics into spontaneous speech. Here is where many current pronunciation textbooks can help, despite their stress-timed orientation, by offering interactive exercises to engage learners in authentic, purpose-driven talk that will call on the rhythm practice they have done.

APPROPRIATE CONTENT FOR ANCHOR PLACEMENT

Ideally our pronunciation instruction should match learners' English proficiency. This ideal raises the ever-present tension between accuracy and simplicity. Teaching experience and data inform our advice on anchor placement.

For beginning adult learners, I recommend Wells' anchor-placement rule: "Only the *first* content word receives an accent" (2006, p. 207 Wells' emphasis). By content word he means noun, adjective, verb, and adverb. To these he adds information-question words, negative words, and demonstrative subject pronouns (pp. 237-40). The first of these in a phrase is the anchor. Otherwise, there is no anchor.

I make this recommendation advisedly. Instructors should understand the consequences of this choice. Wells' simplification misplaces the anchor in certain phrases beginning with adjectives, verbs, adverbs, and indefinite subject pronouns.

Initial Adjectives and Verbs

His popular seminar is rarely canceled. I live seven blocks from campus.

ADJECTIVE + #NAIL + PRIMARY VERB + #NAIL + PRIMARY

In the first case, Wells' rule puts an anchor on *popular* instead of *seminar,* unexpectedly emphasizing the very adjective that native speakers skip over when a later #NAIL choice is available.[8] In the second, his rule predicts the anchor on *live* instead of *seven,* the same verb that native speakers bypass when a #NAIL word occurs later in the phrase.[9]

Initial Adverbs

Pike (Fries 1945:64) found that adverbs of degree modifying adjectives and adverbs are deaccented so regularly that he categorized them with function words. Our data leads to a similar conclusion, with this refinement: Deaccenting is uniform only with the most common adverbs, e.g., *fairly, kind of, pretty, quite, rather, really, so, somewhat, sort of, too, very,* but does not extend to adverbs of degrees that are inherently emphatic such as *awfully, utterly, thoroughly, absolutely, completely, terribly, extremely.*

By contrast, Wells treats all adverbs alike, consistent with his general rule. *Pretty, quite, rather, really, so, very* in anchor position are accented in his materials; common adverbs of degree therefore sound emphatic when they would not be so in everyday speech.

I'm quite relaxed about it. (p. 190) We can be really proud of ourselves. (p. 58)

Initial Indefinite Subject Pronouns

In his most detailed discussion of function words that can carry an anchor, Wells (pp. 237-40) omits indefinite subject pronouns, although he uses *all, anyone, both, everyone, nothing, others, some, someone, something* as subjects elsewhere in his text. Pike lists these as a category of content words; we also include them as one of the four groups of loud function words. By omitting indefinite subject pronouns as accentable, learners will misplace the anchor in phrases similar to these constructed examples:

---

[8] Our analysis of more than 6800 phrases in his text (and available audio) shows that Wells (2006) does not have a single instance of this structure: Adjective + #NAIL + primary peak. It is problematic for his first-content-word rule.

[9] In the same sample referred to in note 8, Wells (2006) has only one marked instance of the form Verb + #NAIL + primary peak, *I'd like a pound of apples, please* (p. 146). He puts an anchor on *pound*, breaking his own rule.

Everyone tried to recite the pledge. Some were unfamiliar with the words.

Given these errors, should Wells' simplification be recommended at all? The reality is that language systems are complex; we teach beginners inaccuracies whenever we simplify language descriptions. The challenge is to assess the impact: Does the simplification capture the essence of a behavior most of the time despite the errors? How do mispredictions affect communication? Our field's traditional version of stress timing, for example, yields a non-English rhythm nearly *all* the time. The costs to communication are detailed in Hahn (1999) and Dickerson (2016). In stark contrast, Wells' simplification captures the neutral rhythm of native speakers at least 90% of the time, in our estimation, with minor communication damage. Adding value to his rule is that it avoids complexities in the native-speaker rule that can lead some beginners to reject anchor placement altogether. An important byproduct of analyzing mispredictions is that we know exactly what to teach when learners are ready for the next step. This is why we say that, for beginners, Wells' simplification trumps absolute accuracy.

For advanced learners, I recommend teaching the anchor-placement rule that native speakers use (including indefinite subject pronouns as part of loud function words and excluding common adverbs of degree from #NAIL). The anchor-placement protocol correctly predicts a neutral anchor in all four cases above. The benefits are clear: When listeners have high expectations for accuracy, it is important not to surprise them with unexpected emphases nor confuse them with less-than-helpful peaks. Wells (2006, p. 2) tells us why: "[N]ative speakers of English . . . do not make allowances for errors in intonation [including accent placement]. [T]hey assume that—when it comes to intonation—you mean what you say." This is why, for advanced learners, accuracy trumps simplicity.

CONCLUSION

In their move from stress-timed rhythm to the two-peak profile, we advise instructors to resist the urge to toss out the BABY—important truths and pedagogical practices—with the bathwater of stress timing. We also encourage them to keep their favorite stress-timed pronunciation text. To make the transition, however, they do need to be strategic in using the text—insisting on short phrases to simulate spontaneous speech, and ignoring exercises that accent every content word and loud function word and that urge regular timing. These small changes to instruction can make a dramatic difference in their students' spontaneous speech.

ACKNOWLEDGEMENTS

I wish to thank Lonna Dickerson, Yelena Tower, John Levis, Mary O'Brien, and an anonymous reviewer for their helpful feedback during the evolution of this paper.

ABOUT THE AUTHOR

Wayne Dickerson is professor emeritus in the Department of Linguistics at the University of Illinois at Urbana-Champaign where he taught courses in English phonology (online and face-to-face) and ESL pronunciation. His research focuses on pedagogical applications of phonetics and phonology, pronunciation pedagogy, the value of orthography for learners, phonological variability, and pronunciation assessment. 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, (The University of Michigan Press).

Contact information:
Wayne Dickerson
7 Hale Haven Court
Savoy, IL 61874 U.S.A.
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.

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.

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

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 Teaching and Learning Conference* (ISSN 2380-9566), Santa Barbara, CA, September 2014 (pp. 184-196). Ames, IA: Iowa State University. https://apling.engl.iastate.edu/psllt-archive/

Dickerson, W. (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. https://apling.engl.iastate.edu/psllt-archive/

Dickerson, W. & Hahn, L. (forthcoming). *Speechcraft: Discourse pronunciation for advanced learners* (2nd ed.). Ann Arbor: The University of Michigan Press.

Fries, C. (1943). *Intensive course in English for Latin-American students*. Vols. 1-4. Ann Arbor: The University of Michigan Press.

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

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

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

Hahn, L. (1999). Native speaker's reactions to non-native stress in English discourse. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign (UMI No. AAT #9944870).

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

Halliday, M. (1967). *Intonation and grammar in British English.* The Hague: Mouton.

Hewings, M. (2007). *English pronunciation in use – Advanced.* Cambridge: Cambridge University Press.

Kjellin, O. (1999). Accent addition: Prosody and perception facilitates 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.

Murphy, J. & Baker, A. (2015). History of ESL pronunciation teaching. In M. Reed & J. Levis (Eds.). *The handbook of English pronunciation.* Chichester, UK: John Wiley & Sons Ltd.

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

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

Prator, C., Jr. & Robinett, B. (1985). *Manual of American English pronunciation* (4th ed.). New York: Holt, Rinehart and Winston.

Sardegna, V. (2009). Improving English stress through pronunciation learning strategies. Unpublished doctoral dissertation, University of Illinois, Urbana. (UMI Dissertation Publishing, ISBN: 9781243619211).

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

Zawadzki, H. (1994). *In Tempo, an English pronunciation course.* Sydney, NSW: National Center for English Language Teaching and Research, Macquarie University.

Johnson, D.O., Kang, O. (2017). Measures of intelligibility in different varieties of English: Human vs. machine. 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. 58-72). Ames, IA: Iowa State University.

MEASURES OF INTELLIGIBILITY IN DIFFERENT VARIETIES OF ENGLISH: HUMAN VS. MACHINE

David O. Johnson, University of Kansas
Okim Kang, Northern Arizona University

This paper demonstrates the feasibility of a tool for measuring the intelligibility of English speech utilizing an automated speech system. The system was tested with eighteen speakers from countries representing six Englishes (American, British, Indian, South African, Chinese, and Spanish) who were carefully selected to represent a range of intelligibility. Intelligibility was measured via two different methods: transcription and nonsense. A computer model developed for automated oral proficiency scoring based on suprasegmental measures was adapted to predict intelligibility scores. The Pearson's correlation between the human assessed and computer predicted scores was 0.819 for the nonsense construct and 0.760 for the transcription construct. The inter-rater reliability Cronbach's alpha for the nonsense intelligibility scores was 0.956 and 0.932 for the transcription scores. Depending on the type of intelligibility measure, the computer utilized different suprasegmental measures to predict the score. The computer employed 11 measures for the nonsense intelligibility score and eight for the transcription score. Only two features were common to both constructs. These results can lead L2 researchers to different perspectives of measuring intelligibility in future research.

INTRODUCTION

In L2 pronunciation, the importance of intelligible speech has been emphasized both in classroom and assessment contexts. Researchers have aimed to determine the specific features that affect intelligibility (Field, 2005; Hahn, 2004) and assessment scores for listeners (Iwashita, Brown, McNamara, & O'Hagan, 2008; Kang, 2013). However, there has been no universally accepted method of measuring intelligibility (Munro & Derwing, 1999). Intelligibility has been ascertained most commonly using transcription or other methods (e.g., true/false statements in Munro and Derwing, (1995) or nonsense, or filtered speech (Kang, Thomson, & Moran, 2015). Due to the technical aspects of such measures, their operationalization has often been supplemented by other constructs, i.e., comprehensibility or accentedness. Currently, advances in computing technology and artificial intelligence have produced automated systems (e.g., SpeechRater^{SM}) that can assess oral proficiency of accented speech, but not intelligibility. The advantages of automated systems are that they can be faster, less expensive, more consistent, and equitable scoring.

This paper introduces an exploratory method of providing an alternative and complementary tool for measuring the intelligibility of different varieties of World Englishes using an automated speech system. The system was tested with 90 sentences from a corpus representing the three circles of World Englishes (Kachru, 1992). Eighteen speakers from countries representing six English varieties (American, British, Indian, South African, Chinese, and Spanish) are included. Sixty listeners from the respective countries listened to speech stimuli to determine the intelligibility of the speech using transcriptions and scalar judgments. The computer system used a machine learning model with suprasegmental measures being the input and the output being an intelligibility measure. The model was trained using the annotated World Englishes corpus and tested on the World Englishes corpus with k-fold cross-validation. Correlations were conducted to compare the computer's calculated intelligibility scores with those of humans. We will discuss how different sets of phonological features influenced the computer's determination of the intelligibility of different varieties of English. These results suggest different perspectives for measuring intelligibility that can be applied in future research.

Intelligibility Measures

There is a growing recognition that L2 speech should aim for intelligibility rather than nativeness (Levis, 2005). In light of this trend, various methods have been utilized in measuring intelligibility in the field of L2 pronunciation. Currently the most commonly used method is a transcription test, which requires a participant to listen to a sound file and transcribe it. Intelligibility scores are based on the percentage of an utterance or word that is transcribed correctly by listeners (Derwing & Munro, 1997). A less-frequently-used method is a cloze test that asks listeners to fill in blanks from a transcript of speech (Smith & Nelson, 1985). The number of words correctly identified determines intelligibility scores. Munro and Derwing (1995) also used True/False judgments in which listeners are asked to make true/false decision about a short sentence they hear. This approach assumes that more intelligible speech will allow listeners to correctly understand the intended message and to correctly evaluate the truth or falsity of sentences. More recently, Kang et al. (2015) introduced exploratory methods of measuring intelligibility, i.e., nonsense statement and filtered speech methods. The nonsense statement task involves decontextualized sentences, which do not make sense semantically (e.g., "Our deaf ads traced my ants."). Listeners were asked to type missing content words into blank boxes provided. Each nonsense sentence receives a score based on the number of correct content words.

Overall, even though varied techniques are available for assessing L2 speech intelligibility, how best to measure intelligibility is still not well understood nor are some intelligibility measures necessarily easy to implement. In the current study, we attempted to explore a new way of measuring intelligibility by adopting a computer model developed for automated oral proficiency scoring to predict intelligibility scores. The findings of the study are exploratory and should not be over-generalized to other contexts of speech corpora or language assessment.

METHODS

World Englishes Speech Corpus

The data set we used is from a project investigating intelligibility of different varieties of World Englishes. Because we wanted to do an exploratory experimental study that could simply compare human vs. machine's intelligibility ratings, for the sake of convenience, we used the World Englishes corpus.

The World Englishes speech files were developed as part of a TOEFL listening test project supported by Educational Testing Service. Eighteen speakers (ages 30-50) were carefully chosen, one female and two males from each of six countries: United States and England (Inner

Circle), India and South Africa (Outer Circle), and Mexico and China (Expanding Circle). All of the Inner Circle and Outer Circle speakers were highly proficient in English but retained a noticeable foreign accent. The speakers were selected to represent a range of intelligibility, as determined by nine trained raters' scalar judgments. The speakers were either currently teaching in English as professors or graduate students.

The speakers were asked to record in a quiet location 72 nonsense sentences, 72 true/false sentences, and 18 iBT listening passages (4-5 minutes) to be utilized for the intelligibility test and computer model training. Using Audacity, a research assistant acoustically edited noises and sound quality for practical uniformity and added three seconds of pause time before and after each passage.

The intelligibility of the speakers was scored by 60 listeners, consisting of ten listeners representing each of the six World English varieties. The listeners were recruited both nationally and internationally. Listeners of non-inner circle English varieties were highly proficient with TOEFL iBT scores greater than 100. They were undergraduate and graduate students (43% males and 57% females). The intelligibility scoring was administered via computer, using headphones, and in a highly controlled laboratory setting under careful supervision. The speech files were randomly presented to the listeners. Two measures of intelligibility were used: *transcription* (Derwing & Munro, 1997) and *nonsense* (Kang et al., 2015).

For the transcription measure, listeners heard each of the 18 speakers read four sentences (72 sentences total), four to eight words in length, that were syntactically correct, but semantically incorrect. An example of the sentences is '*gasoline is an excellent drink'*. They listened to each sentence one time only and then were asked to transcribe what was said. Each speaker received an intelligibility score of 0-100% based on the number of words the listeners transcribed correctly in all four sentences. The actual transcription intelligibility scores ranged from 88.02 to 99.14%.

The nonsense intelligibility score was determined by listening to four nonsense sentences recited by each of the 18 speakers. The nonsense sentences were semantically meaningless, though syntactically normal, containing frequently used monosyllabic words. The sentences were adopted from studies on native language (L1) intelligibility (Nye & Gaitenby, 1974; Picheny, Durlach, & Braida, 1985). To score intelligibility, the listeners were asked to type missing words from the nonsense sentences into each of four text boxes. An example nonsense sentence with the missing words underlined is '*The <u>tall</u> <u>kiss</u> can <u>draw</u> with an <u>oak</u>'*. The speaker's nonsense intelligibility score was then calculated as the total number of blanks out of 16 correctly filled in by the listeners. 8.62 to 13.62 was the actual range of the nonsense intelligibility scores. Scores from the both methods were normalized for comparison. Table 1 shows the raw and normalized values for both measures of intelligibility assessed by the humans: nonsense and transcription.

Table 1

Raw and normalized intelligibility scores

Speaker	Gender	World English	Nonsense Raw	Nonsense Normalized	Transcription Raw	Transcription Normalized
1	M	Inner	13.62	6	98.31	6
2	M	Inner	13.17	6	98.65	6
3	F	Inner	12.62	5	98.25	6
4	F	Inner	12.52	5	99.14	6
5	M	Inner	11.47	4	98.93	6
6	M	Inner	11.35	4	99.13	6
7	M	Outer	13.28	6	98.82	6
8	M	Outer	11.70	4	95.91	5
9	M	Outer	10.90	3	98.46	6
10	F	Outer	9.88	2	93.37	3
11	M	Outer	8.75	1	88.02	1
12	F	Outer	8.62	1	92.27	3
13	M	Emerging	13.13	6	94.30	4
14	F	Emerging	12.92	5	97.91	5
15	M	Emerging	10.45	3	95.80	4
16	F	Emerging	10.33	3	96.85	5
17	M	Emerging	10.12	3	92.62	3
18	M	Emerging	9.10	1	96.71	5

Intelligibility Score: Computer Prediction

Figure 1 illustrates the computer model employed to predict the intelligibility scores.

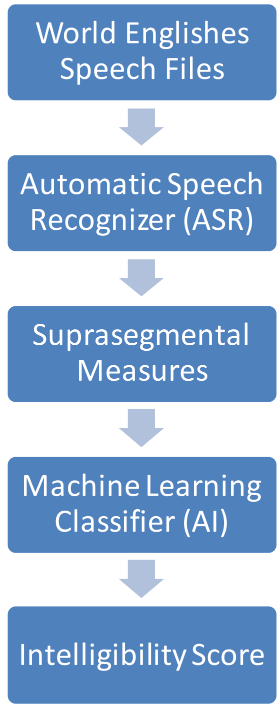


Figure 1. Computer model for predicting intelligibility scores

To predict the intelligibility scores, the computer first analyzed the 144 speech files with an automatic speech recognizer (ASR). From the output of the ASR, the computer calculated 35 suprasegmental measures of rate (e.g., syllables per second, articulation, phonation ratio), pause (e.g., silent and filled pauses per minute and mean length), stress (e.g., pace, space, percent tone units with termination), pitch (e.g., pitch range, mean prominent syllable pitch), paratone (e.g., low terminations followed by high key resets), and intonation (e.g., percent of tone choice and relative pitch). Then, the computer utilized AI (artificial intelligence) machine learning techniques to predict normalized intelligibility scores from one to six. The computer utilized a genetic algorithm to select the most salient suprasegmental measures and then built a decision tree classifier to predict the intelligibility scores from salient suprasegmental measures. The classifier was trained to achieve the best human-computer correlation by 3-fold cross-validation of the speech files. Each set of the 72 speech files was used to train a separate computer model for the two different intelligibility measures: transcription and nonsense. The computer model was developed for automated oral proficiency scoring (Johnson, Kang, & Ghanem, 2015), but in this case the output was the intelligibility score instead of the proficiency score.

RESULTS

First, we examined the Pearson's correlation between the intelligibility scores assessed by the humans and those predicted by the computer model. Figure 2 gives the Pearson's correlation between the human assessed intelligibility scores and the computer predicted scores. The correlation between the human and computer nonsense intelligibility scores is 0.819; the correlation between the two transcription intelligibility scores is 0.760.

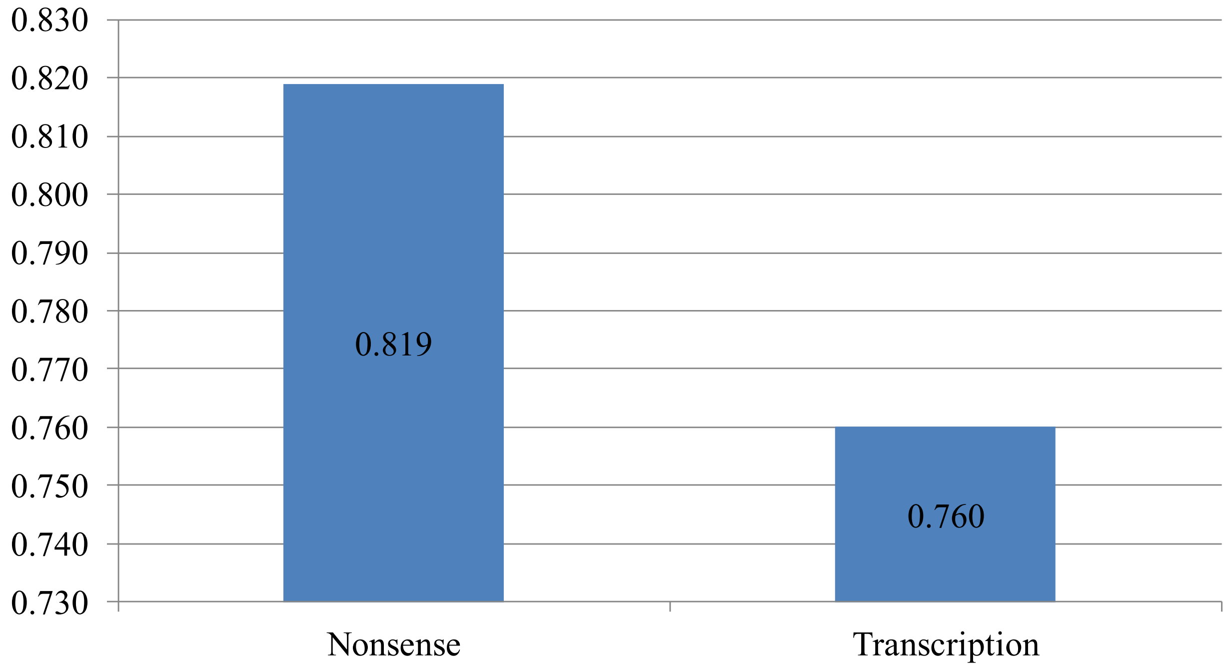


Figure 2. Human-computer correlation (*r*)

Next, we analyzed the individual correlations between humans for the nonsense intelligibility score as depicted in Figure 3. The pair-wise correlations between the 60 humans ranged from 0.989 to -0.581.

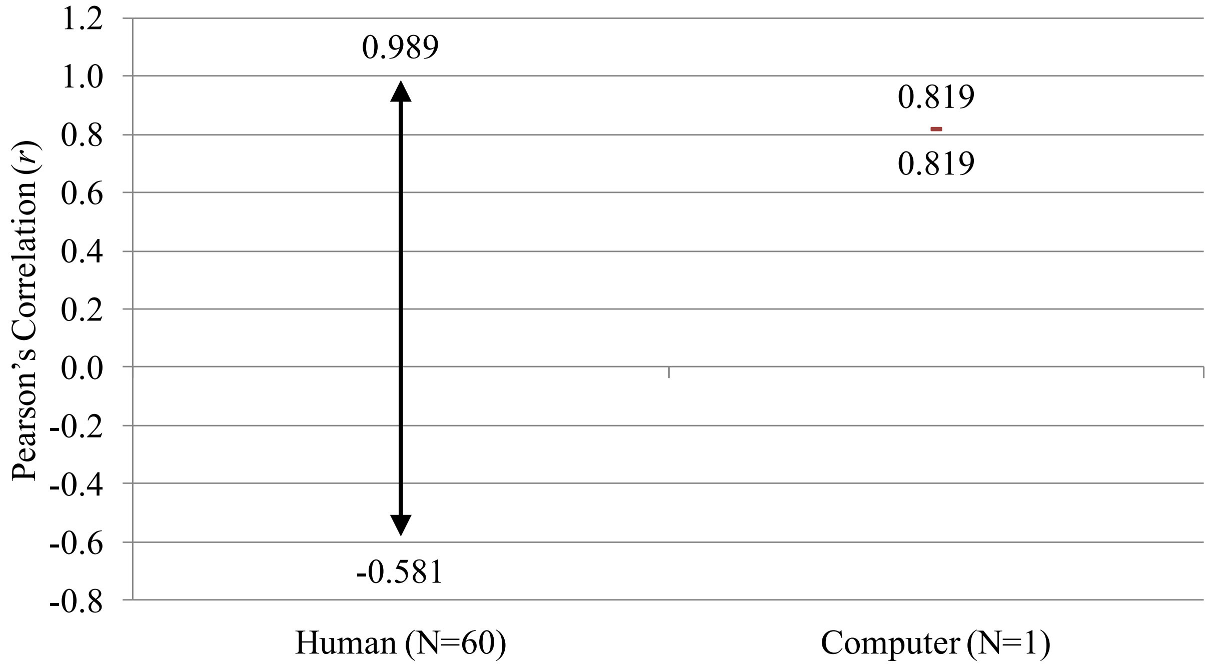


Figure 3. Human-human correlation vs. human-computer correlation (r) for nonsense scores

Next, we investigated interrater reliability with Cronbach's alpha. Figure 4 shows that the Cronbach's alpha (α=0.85) for the nonsense intelligibility scores was 0.956 (N=60) and 0.710 (N=5).

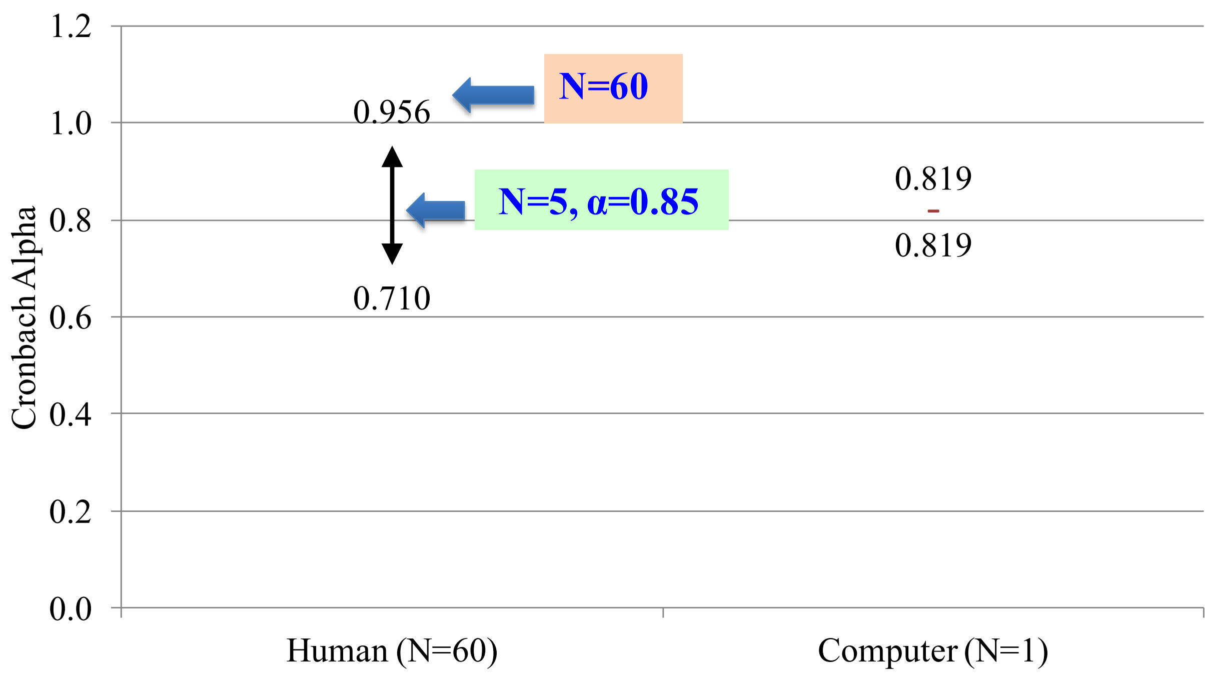


Figure 4. Interrater reliability for nonsense intelligibility scores

As described in Figure 5, the Cronbach's alpha for the transcription intelligibility scores was 0.932.

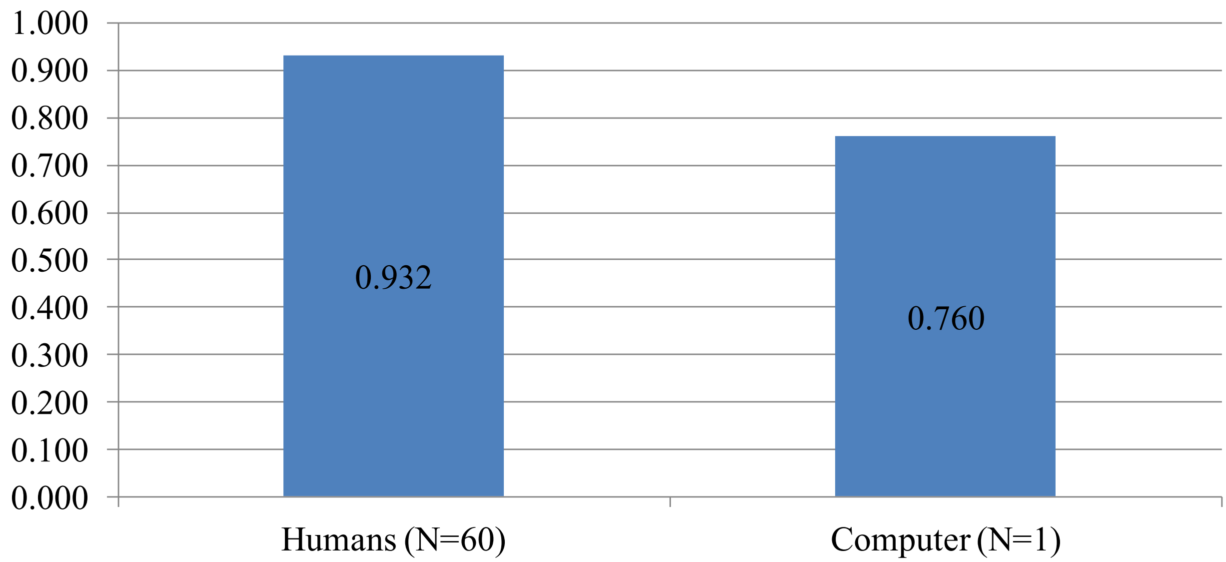


Figure 5. Inter-rater reliability for transcription intelligibility scores

After comparing the intelligibility scores appraised by the humans and those calculated by the computer model, we looked at what salient suprasegmental features were selected by the computer model. Table 2 gives the salient suprasegmental features selected by the genetic algorithm of the computer model to forecast each of the two intelligibility scores (marked with an X).

Table 2

Salient suprasegmentals in the computer model

Type	Suprasegmental	Nonsense	Transcription
Rate	phonation time ratio		X
	syllables per second		X
Pause	mean length of filled pauses	X	
	mean length of silent pauses	X	
	number of silent pauses per minute	X	
Stress	number of prominent syllables per run (pace)	X	
	prominence characteristics		X
	proportion of prominent syllables (space)		X

Pitch	average pitch of new prominent syllables	X	
	overall pitch range	X	X
	average height of terminating pitch	X	
Tone choice	% mid falling tone choices	X	
	% high falling tone choices	X	
	% mid fall-rise tone choices		X
	% high fall-rise tone choices	X	
	% low rising tone choices	X	X
	% mid rising tone choices		X

We found that, depending on the type of intelligibility measure, the computer picked different features. For the nonsense intelligibility score, the computer employed 11 features: mean length of filled pauses, mean length of silent pauses, number of silent pauses per minute, number of prominent syllables per run (pace), average pitch of new prominent syllables, overall pitch range, average height of terminating pitch, % falling mid tone choices, % falling high tone choices, % fall-rise high tone choices, and % rising low tone choices. On the other hand, the computer utilized only eight features for the transcription intelligibility score: phonation time ratio, syllables per second, prominence characteristics, proportion of prominent syllables (space), overall pitch range, % fall-rise mid tone choices, % rising low tone choices, and % rising mid tone choices. Only two features, overall pitch range and % rising low tone choices, were common to both nonsense and intelligibility predictions.

DISCUSSION

The results show there is a correlation between the human assessed and computer predicted scores for both intelligibility constructs, nonsense ($r = 0.819$) and transcription ($r = 0.760$). The computer utilized different suprasegmental measures to predict the score for each construct, 11 measures for nonsense and eight for transcription, with two the same for both constructs (i.e., pitch range and mid rising). In comparing this work with other similar research, Kang et al. (2015) found mean length of run, expected pause ratio, number of prominent syllables per run, word stress errors, % of falling tone choice, and vowel and consonant substitution/deletion errors as the salient variables in human ratings when the nonsense sentence was utilized to gauge intelligibility. Table 3 compares these with the ones selected by the computer model in this study.

Table 3

Salient pronunciation features that predict intelligibility with the nonsense sentence construct

Human Ratings (Kang et al., 2015)	Computer Modeling
mean length of run	
expected pause ratio	mean length of silent and filled pauses number of silent pauses
number of prominent syllables per run word stress errors	number of prominent syllables per run
	overall pitch range
% of falling tone choice	% of falling tone choice % of rising tone choice
vowel & consonant substitution/deletion errors	The computer model is only trained for suprasegmental features at the moment.

The salient variables for the human and computer ratings for the nonsense sentence intelligibility measure differ in four areas: pausing, prominence, pitch, and tone choice. (Note: the computer model was only trained for suprasegmental measures; therefore, any discussion of segmentals is excluded). For pausing, the salient variables in human ratings were mean length of run and expected pause ratio. One could argue that these are very similar to the measures selected by the computer because they gauge how a speaker uses pausing (both filled and silent) to articulate the speech. Thus, from this perspective both humans and the computer model found pausing to be a salient predictor of intelligibility. Both discourse and intonation units are delineated by pauses (Brazil, 1997; Wagner & Watson, 2010). Previous studies have recognized that non-native speakers pause silently more frequently, longer, and more unevenly than native speakers (Anderson-Hsieh & Venkatagiri, 1994; Riggenbach, 1991; Rounds, 1987).

With regard to prominence, pace (number of prominent syllables per run) was a salient variable for humans and the computer model. The computer did not measure word stress errors; accordingly, this variable could not be compared. It appears that both humans and the computer model agree that prominence significantly predicts the intelligibility of accented speech. The prominent syllable is a basic aspect of Brazil's (1997) model of English prosody. The proper use of prominent syllables by speakers should then be an important suprasegmental measure of intelligibility. Thus, it is consistent that the computer model employed number of prominent syllables per run as a predictor of intelligibility.

The computer model also used average pitch of new prominent syllables, overall pitch range, and average height of terminating pitch as predictors of intelligibility. The saliency of average pitch of new prominent syllables and average height of terminating pitch is suspect because these are paratone measures and the nonsense test did not really include paratones. Thus, more research is necessary to determine if these are salient or an anomaly. Likewise, overall pitch range is more likely an indication of the variation in voice pitch between speakers, rather than an indication of an individual speaker utilizing intonation as a discourse signal to begin, maintain, and end a thought group or as a means of differentiating the information content of specific lexical items (Cutler, Dahan, & Donselaar, 1997; Kang, Rubin, & Pickering, 2010).

In the category of tone choice, the computer model and the human study mutually found falling tone choices to be indicative of relative intelligibility. This is in harmony with earlier research which puts forward that lower proficiency speakers tend to overuse falling tones until they learn that it has a negative impact on intelligibility (Kang, 2012). Additionally, the computer model found the use of rising tone to be a significant predictor of NNS intelligibility. That is, intelligible speech tends to contain more use of rising tone but less use of falling tone. Overall, when employing the nonsense intelligibility test, the human raters and computer model agreed that pausing, prominence, and tone choice are salient features of intelligibility.

Mean length of filled pauses is the only one of the computer model's predictors not supported by other research. According to Goldman-Eisler (1968), filled pauses may imply more regarding a speaker's style of articulation and cognitive load and less about speaking proficiency. Similarly, Fulcher (1996) said that more capable speakers make an impression on listeners because they pause for distinct reasons, not because they pause at a different rate than less capable speakers.

Kang et al. (2015) also found mean length of run, syllable per second, proportion of prominent syllables (space), word stress error, % of rising tone choice, vowel consonant substitution/deletion errors, consonant deletions, syllable reductions, and consonant cluster errors to be the salient variables when assessing intelligibility with the transcription construct. These are contrasted with those utilized by our computer model in Table 4.

Table 4

Salient pronunciation features that predict intelligibility with the transcription construct

Human Ratings (Kang et al., 2015)	Computer Modeling
mean length of run syllable per second	phonation time ratio syllables per second
proportion of prominent syllables (space) word stress error	proportion of prominent syllables (space) prominence characteristics
	overall pitch range
% of rising tone choice	% of rising tone choice
vowel consonant substitution/deletion errors & consonant deletions, syllable reductions, consonant cluster errors	The computer model is only trained for suprasegmental features at the moment.

Like the nonsense construct, except for overall pitch range, the human study and the computer model seem to agree on the salient features for measuring intelligibility by the transcription method. They both found prominence (prominence characteristics and space), falling tone choice (% fall-rise mid tone choices), and rising tone choice (% rising low and mid tone choices) to be prognosticators of intelligibility, all of which are consistent with prior research as discussed above.

In the area of speech rate, both concur on syllables per second as salient. This is in line with the conjecture of Kormos and Dénes (2004) that proficiency is a speech rate phenomenon in addition to an intonational one. Ginther, Dimova, and Yang (2010) noted strong to moderate correlations linking oral English proficiency scores and speech rate (i.e., syllables per second or syllable rate), articulation rate, and mean length of run (i.e., the average number of syllables per run). The computer also found that speech rate and phonation time ratio, which is speech rate divided by articulation, were predictive of intelligibility. Mean length of run was a leading measure in the human study. However, the computer did not find mean length of run to be an indicator of intelligibility. This may be because of the study's use of sentences rather than extended discourse. Even though the computer did not end up using it in its scoring, one could argue that speech rate, articulation rate, and mean length of run are inter-related and it is not necessary to consider all three of them.

With regard to prominence, the computer model is consistent with the human study in the salience of space as a predictor of intelligibility. The human study also found word stress error to be salient and the computer model found prominence characteristics to be. The human study did not measure prominence characteristics, nor did the computer model consider word stress errors, so it is impossible to compare these two aspects. Hence, rate, stress, and tone choice emerge as salient features in both the computer model and human study when the measuring intelligibility via the transcript construct.

CONCLUSION

This paper presents an exploratory approach to automating intelligibility scoring using a computer model that predicts scores based on suprasegmental measures derived from an automated speech system. The results suggest the importance of suprasegmentals in human judgments as well as machine scoring. We also found that the salient suprasegmental measures used by the computer model depend on which intelligibility measurement method is employed, either nonsense or transcription. Further research is needed on this topic, however. First, additional work is necessary to improve the accuracy of our current computer modeling. The accuracy of the underlying components of the suprasegmental measures (e.g., prominent syllables) obtained from the output of the ASR varies due to the inherent error rates of the instrumentation, algorithms, and machine learning techniques applied. Second, the current computer model only made use of suprasegmental features. Incorporating segmental features along with other linguistic properties (e.g., grammatical and lexical features) into the computer model could improve its prognostic capabilities.

Currently, the computer model predicts a normalized intelligibility score ranging from one to six. This could be expanded to the full range of the intelligibility measures which is 0-100 for the transcription score and 0-16 for the nonsense score. The model could also be enhanced to predict intelligibility based on other assessment constructs such as accentedness or comprehensibility. A larger corpus of speakers is also recommended to validate the computer model. Although the corpus was carefully created to represent a wide range of World Englishes speakers, it only provided a training set of 12 speakers and a testing set of 6 speakers.

ABOUT THE AUTHORS

Okim Kang is an Associate Professor in the Applied Linguistics Program at Northern Arizona University, Flagstaff, AZ, USA. Her research interests are speech production and perception, L2 pronunciation and intelligibility, L2 oral assessment and testing, automated scoring and speech recognition, World Englishes, and language attitude.
Author's contact information: okim.kang@nau.edu

David O. Johnson is a Lecturer in the Electrical Engineering and Computer Science department at the University of Kansas in Lawrence, KS, USA. At the time of this research, he was a post-doctoral researcher in the Applied Linguistics Speech Laboratory at Northern Arizona University, Flagstaff, AZ, USA developing software and computer models to automatically score English language proficiency and intelligibility. He received his BSEE and MSEE from Kansas State University and his PhD in Computer Science from the University of Kansas. Prior to a post-doctoral research appointment at the Eindhoven University of Technology in the Netherlands, he was an Adjunct Professor in the Computer Science Electrical Engineering

department at the University of Missouri – Kansas City. He is interested in natural language processing and human-robot interaction.
Author's contact information: Email: davidojohnson@aol.com

REFERENCES

Anderson-Hsieh, J., & Venkatagiri, H. (1994). Syllable duration and pausing in the speech of Chinese ESL speakers. *TESOL Quarterly*, 28, 807–812.

Brazil, D. (1997). *The communicative value of intonation in English*. Cambridge: Cambridge University Press.

Cutler, A., Dahan, D., & Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech*, 40, 141–201.

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

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

Fulcher, G. (1996). Does thick description lead to smart tests? A data-based approach to rating scale construction. *Language Testing*, 13, 208–238.

Ginther, A., Dimova, S., & Yang, R. (2010). Conceptual and empirical relationships between temporal measures of fluency and oral English proficiency with implications for automated scoring. *Language Testing*, 27(3), 379-399.

Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech*. London: Academic Press.

Hahn, L.D. (2004). Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly, 38*, 201–223. doi: 10.2307/3588378.

Iwashita, N., Brown, A., McNamara, T., & O'Hagan, S. (2008). Assessed levels of second language speaking proficiency: How distinct? *Applied Linguistics*, *29*(1), 24-49.

Johnson, D. O., Kang, O., & Ghanem, R. (2015). Language proficiency ratings: human vs. machine. In *J. Levis, H. Le, I. Lucic, E. Simpson, & S. Vo (Eds). Proceedings of the 7th Pronunciation in Second Language Learning and Teaching Conference*, (pp. 119-129), ISSN 2380-9566, Dallas, TX, October 2015. Ames, IA: Iowa State University.

Kachru, B. B. (1992). *The other tongue: English across cultures*. University of Illinois Press.

Kang, O. (2012). Impact of rater characteristics and prosodic features of speaker accentedness on ratings of international teaching assistants' oral performance. *Language Assessment Quarterly*, 9(3), 249-269.

Kang, O. (2013). Relative impact of pronunciation features on ratings of non-native speakers' oral proficiency. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference* (pp. 10-15). Ames, IA: Iowa State University.

Kang, O., Thomson, R., & Moran, M. (2015), Intelligibility of different varieties of English: The effects of incorporating "accented" English into high stakes assessment. Presentation at *American Association of Applied Linguistics Conference*, Toronto, ON, Canada, March 21–24, 2015.

Kang, O., Rubin, D., & Pickering, L. (2010). Suprasegmental measures of accentedness and judgments of language learner proficiency in oral English. *The Modern Language Journal, 94*(4), 554-566.

Kormos, J., & Denes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, 32, 145–164.

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

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. (1999). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 49(s1), 285-310.

Nye, P. W., & Gaitenby, J. H. (1974). The intelligibility of synthetic monosyllabic words in short, syntactically normal sentences. *Haskins Laboratories Status Report on Speech Research*, 37(38), 169-190.

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1985). Speaking clearly for the hard of hearing intelligibility differences between clear and conversational speech. *Journal of Speech, Language, and Hearing Research*, 28(1), 96-103.

Riggenbach, H. (1991). Towards an understanding of fluency: A microanalysis of nonnative speaker conversation. *Discourse Processes*, 14, 423–441.

Rounds, P. (1987). Characterizing successful classroom discourse for NNS teaching assistant training. *TESOL Quarterly*, 21, 643–672.

Smith, L. E., & Nelson, C. L. (1985). International intelligibility of English: Directions and resources. *World Englishes*, 4(3), 333-342.

Wagner, M., & Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes*, 25(7-9), 905-945.

Koffi, E. (2017) An acoustic phonetic account of the confustion between [ɹ] and [l] in seven varities of L2 English: Focus on intelligibility and accentedness. 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. 73-84). Ames, IA: Iowa State University.

AN ACOUSTIC PHONETIC ACCOUNT OF THE CONFUSION BETWEEN [ɹ] AND [l] IN SEVEN VARIETIES OF L2 ENGLISH: FOCUS ON INTELLIGIBILITY AND ACCENTEDNESS

Ettien Koffi, St. Cloud State University

The lateral [l] and the rhotic [ɹ] are classified as liquids because their places and manners of articulation overlap in many languages (Ladefoged & Maddieson 1996, p.185). As a result, when some L2 speakers of English produce them, these two segments are perceptually indistinguishable to some speakers of English. This is likely to cause unintelligibility for hearers of General American English (GAE). Other L2 speakers produce them clearly and distinctly. Intelligibility is not compromised, yet their speech is perceived as heavily accented. Both pronunciation patterns are investigated acoustically to determine thresholds at which unintelligibility or accentedness is perceived. The participants in this study are 10 Arabic, 10 Japanese, 10 Korean, 10 Mandarin, 11 Slavic, 6 Somali, and 10 Spanish L2 speakers. Their pronunciations of [l] and [ɹ] are compared and contrasted with those produced by 10 native speakers of American English. All 77 participants read the same *Speech Accent Archive* text. The acoustic correlates studied in this paper are F3 and vibration rates. Catford's (1987) relative functional load (RFL) data is used to determine the rate of intelligibility, while Ladefoged (1996)'s trilling threshold is used to gauge the degree of accentedness of the [ɹ]s produced by non-native speakers.

INTRODUCTION

Lindau (1985, pp. 157-168) wrote a paper entitled *The Story of /r/* in which she investigated the phonetic properties of /r/ in four European and seven non-European languages. I follow her methodology but focus instead on the story of /r/ in seven varieties of L2- English. This paper is divided into four sections. The first tabulates the confusion errors resulting from the substitutions of [l] by [ɹ] and vice versa. The second provides acoustic measurements used to possibly account for the confusion. The third discusses the sociolinguistic significance of the confusion. The fourth suggests kinematic exercises that can help improve intelligibility on the one hand, and reduce strongly-accented [ɹ]s on the other hand.

QUANTITATIVE ANALYSIS OF THE CONFUSION BETWEEN /l/ AND /r/

Seventy-seven participants were examined in this study. Ten are native speakers of General American English (GAE). The remaining 67 participants include 10 speakers of Arabic, 10 Japanese speakers, 10 Korean speakers, 10 Mandarin speakers, 11 speakers of South Slavic languages/dialects (5 Croatians and 6 Serbians), 6 Somali speakers, and 10 Spanish speakers. All of them read the same text from the George Mason University's Speech Accent Archive. The text has 16 words containing the grapheme <r>. Eight of them occur prevocalically, seven of which were clusters: <b**r**ing, f**r**om, f**r**esh, **bro**ther, f**r**og, th**r**ee, **r**ed, t**r**ain> and eight others postvocalically: <he**r**, he**r**, sto**r**e, fo**r**, he**r**, broth**er**, fo**r**, he**r**>. The text also has five prevocalic <l>s: <p**l**ease, s**l**abs, b**l**ue, p**l**astic, Ste**ll**a> and four postvocalic <l>s: <ca**ll**, a**l**so, sma**ll**, wi**ll**>. Together the 67 non-native speakers attempted 1,072 [ɹ]s and 603 [l]s, for a total of 1,675 liquids. The quantitative analysis focuses on the 871 prevocalic laterals [l] and rhotics [ɹ]. The remaining 804 segments are not discussed because they occur in syllable codas. It is well known that in this position, liquids are often vocalized or even deleted (Lindau, 1985, p. 157 and Ash, 1982). The quantitative analysis of the narrow phonetic transcription shows that 833 liquids were produced accurately. In 33 instances [l] and [ɹ] were confused with each other. In five other instances, they were misperceived as [n] or [w], as shown in Table 1.

Table 1

Confusion of Matrix of /l/ and /r/

	Perceived Stimuli					
Spoken		[l]	[n]	[ɹ]	[w]	Total
Stimuli	[l]	309	3	23	0	335
	[ɹ]	9	0	525	2	536

The total erroneous pronunciations amount to 4.36%, but the actual confusion errors attributable to prevocalic [l] and [ɹ] represent 3.78% of the data. A closer examination of the data in Table 2 shows that Japanese, Mandarin, and Korean talkers made most of the errors:

Table 2

Confusion Rate by Language Group

	Confusion Rate	**Percentage**
Arabic	0 out of 130	0
Japanese	23 out of 130	17.69%
Korean	6 out of 130	4.61%
Mandarin	7 out of 130	5.38%[1]
Slavic	1 out of 143	.69
Somali	0 out of 78	0
Spanish	0 out of 130	0
GAE	0 out of 130	0

Speakers from these three language groups produced 37 of the 38 errors (97.36%). The Japanese talkers alone made 23 errors (60.52%), followed by Mandarin talkers who made seven errors (18.42%), and by Korean speakers who made six errors (15.78%). However, Mandarin speakers confused [l] and [ɹ] only twice. The five remaining errors have to do with three instances when [l] was transcribed as [n], and twice when [ɹ] was perceived as [w]. Since [l] was confused with [ɹ] only twice, we will consider it an incidental error. We will focus the rest of the paper on the errors made by Japanese and Korean speakers.

FEATURE DISTRIBUTION OF /l/ AND /r/

Why did GMU transcriptionists transcribe [l] and [ɹ] as they did in Japanese and Korean-accented English? Why didn't they make similar perception errors with the other accented Englishes? The answers may lie in the information presented in Table 3 regarding the kinds of liquids found in the native languages of the speakers. The data in the second and third columns are taken from Ladefoged and Maddieson (1996, pp.185, 216). That in the last column is based on information from the *Handbook of the International Phonetic Association* (1999) and Maddieson (1984).

[1] The higher confusion rate of Mandarin-accented English requires some explanations. There were only two instances where [l] and [r] were confused with each other. However, in three instances, [l] was mispronounced as [n], and in two others, [r] was produced as [w]. The actual [l] and [r] confusion rate is 1.53%. Since this paper focuses narrowly on [l] and [r], we will not address the confusion between [l] and [n], nor will we discuss the one between [r] and [w].

Table 3

Articulatory Descriptions of /l/ and /r/

N0	Articulatory Descriptions	IPA Symbols	Languages
1.	Voiced alveolar lateral	l	Mandarin, Slavic, Somali
2.	Voiced laminal dental lateral	l̪	Spanish
3.	Voiced apical alveolar lateral	l̺	Arabic
4.	Voiced apical post-alveolar	l̩	
5.	Voiced laminal post-alveolar	l̠	
6.	Voiced lateral alveolar flap	ɭ	Korean
7.	Voiced alveolar or alveolar trill	r	Arabic, Mandarin, Slavic, Somali, Spanish
8.	Voiced dental or alveolar tap or flap	ɾ	GAE, Korean
9.	Voiced dental or alveolar approximant	ɹ	GAE, Spanish
10.	Voiced post-alveolar flap	ɽ	Japanese
11.	Voiced post-alveolar approximant	ɻ	
12.	Voiced uvular trill	ʀ	
13.	Voiced uvular approximant	ʁ	
14.	Voiced dental or alveolar lateral flap	ɺ	

Two rhotics are listed for Spanish because it has both an approximant flapped [ɹ] and a trilled [r] (Lindau 1985, p.161). Korean also has two liquids. The lateral /ɭ/ is the basic phoneme, but it has [ɾ] as its allophone in intervocalic positions (Lee 1999, pp.120-123). The status of /r/ in Mandarin is in limbo. Some take it to be an allophone of /l/, but the existence of minimal pairs between [l] and [ɹ] (Smith 2010, pp. 88-89) suggests that [ɹ] may be a phoneme in its own right. Okada (1999, p. 118) indicates that the Japanese postalveolar rhotic [ɽ] occurs mainly between vowels. Arabic, Slavic (Croatian and Serbian), and Somali have only trilled [r]s. Except for Japanese, all the languages have a lateral segment.[2] The distributional restrictions on liquids in Japanese and Korean may explain why the pronunciations of [l] and [ɹ] were transcribed with the other sound more than the other groups of L2 English speakers. However, since our interest in this paper is not phonology, we turn our attention to acoustic phonetics for explanations.

ACOUSTIC PHONETIC ACCOUNTS

Before providing the relevant acoustic measurements on which the explanations of the analyses will be based, we must discuss briefly the methodology and equipment used. The recordings were downloaded from the Speech Accent Archive website. A commercial software package called "Wav Pad" was used to convert the videos into audio files and save them as .wav files. The words <Stella> and <red> were selected to study the acoustic characteristics of [l] and [ɹ] produced by all 77 participants. Each file was annotated in Praat (Boersman and Weenink 2012, Version 5.3.16) using Ryan's (2005) Grid.maker.praat. Five tiers were created to study F1, F2, F3, Intensity, and Duration. The spectrographs of <red> produced by MN 143M illustrates the procedure that was followed for all the speakers:

[2] There are disagreements on the IPA symbols used for Japanese and Korean liquids. Some use [ɽ] or [ɺ] for Japanese. Others use [ɽ] or [ɾ] for Korean rhotics.

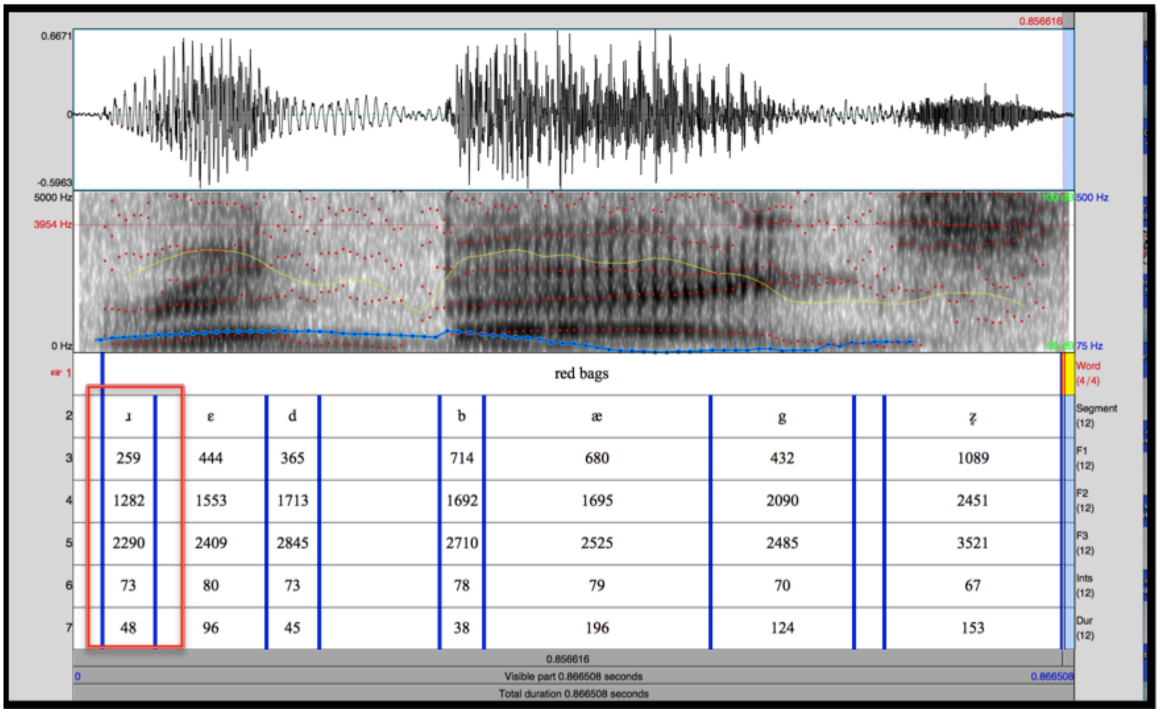


Figure 1. Spectrograph of <red> in <three red bags> by MN 143M

Yoon's (2008) script, Stress-analysis.praat, was used to collect all measurements. A rectangle has been drawn around [ɹ] to draw attention to it since all the other measurements in the spectrogram are irrelevant to this analysis.

F3 Measurements of L2-accented [l]s and [ɹ]s

The data was compiled for each of the 77 participants. The measurements were then averaged per language group, as shown in Table 4.

Table 4.

F3 of [l] and [ɹ]

	F3 of [l]	**F3 of [ɹ]**
Arabic	2,619	2,397
Japanese	2,779	2,674
Korean	2,754	2,697
Mandarin	2,708	2,451
Slavic	2,738	2,517
Somali	2,943	2,483
Spanish	2,869	2,495
GAE	2,827	2,225

The acoustic correlate that phoneticians deem the most robust for discriminating between [l] and [ɹ] is F3 (Lindau 1985, p. 158). It goes without saying that it is the correlate used in this study. Espy-Wilson's (1992:739, 747) study of [l] and [ɹ] in GAE is considered the most authoritative to date. She measured their occurrences in three phonological environments: prevocalic, intervocalic, and postvocalic. The mean F3 measurements for [l] are, respectively, 2,553 Hz, 2640 Hz, and 2630 Hz. The mean in all three positions is 2,607 Hz. The measurements for [ɹ] are as follows: 1,779 Hz prevocalically, 1,720 Hz intervocalically, and 1,830 Hz postvocalically. Its mean F3 in all positions is 1,776 Hz. These measurements are in line with other studies. Generally, the F3 range for [l] is from 2,600 to 3,000 Hz, while that for [ɹ] goes from 1,200 Hz to 2,000 Hz.[3] Ladefoged and Maddieson (1996, p.244) explain that "a lowered third formant is a well-justified specification for the various articulations of the American ɹ."

Cross-linguistic studies of [l] and [ɹ] show similar measurements. Except for apical/dental trills, rarely does one find a rhotic whose F3 is 2,600 Hz. This suggests that any liquid whose F3 reaches 2,600 Hz is very likely to be categorized as a lateral. On the basis of available data on the F3 of liquids, Koffi (2016) has proposed the Liquid Intelligibility Criterion (LIC):

Liquid Intelligibility Criterion (LIC)

[ɹ] masks [l] if its F3 is ≥ 2,600 Hz, unless it is trilled.

[3] The mean of all GAE talkers in Koffi (2016) is 2,225 Hz. This is likely due to the fact that all the [r]s in the data occur before the front vowel [ɛ].

The criterion implies that [l] with F3 less than 2,600 Hz may also be confused with [ɹ]s. However, F3 alone is not a necessary and sufficient condition to predict that [ɹ] and [l] will mask each other. The escape clause "unless it is trilled" makes it clear that trilling is the difference. Ladefoged and Maddieson (1996, p.203) have reported that there is no known lateral trill in any human language. Consequently, any liquid that is trilled is automatically perceived as a rhotic. Six Japanese talkers: Japanese 2F, 4M, 9M, 11F, 12M, 13M, and seven Koreans: 1M, 2F, 4F, 8F, 9M, 10M, 12M, 13F, produced [ɹ]s with F3 values at or beyond 2,600 Hz.[4] Did they trill them? If they did, no confusion would arise. If they did not, confusion is predicted to occur.

Vibration Rates of L2-Accented [l]s and [ɹ]s

Before answering the question from the previous section, let's see how trilling is calculated. According to Lindau (1985, p.166), the determination that a rhotic is trilled or flapped depends on its total duration. Ladefoged (1996, pp. 114-116) provides the following formula for calculating trilling:

$$\text{Vibration Rate in Hz} = \frac{\text{Duration in milliseconds}}{\text{Segment duration Ref}}$$

A rhotic is perceived as trilled if its vibration rate is at or exceeds 22 Hz (Ladefoged and Maddieson 1996, p. 218, 226). This is an important threshold that can be used in tandem with F3 measurements to explain why GMU transcriptionists perceived some Japanese and Korean-accented [l]s as [ɹ]s, and vice versa. The data in Table 5 provides some answers:

[4] The explanations for these measurements found in Koffi (2016, pp. 258-264).

Table 5

Duration and Vibration Rates of [l] and [ɹ]

	Duration in Ms		**Vibration in Hz**		**Difference**
Words	**Stella**	**red**	**Stella**	**red**	**Red - Stella**
Segments	**[l]**	**[ɹ]**	**[l]**	**[ɹ]**	**[ɹ]-[l]**
Arabic Mean	73	40.9	13.59 Hz	24.44 Hz	10.85
Japanese Mean	73.35	70.6	13.63 Hz	14.16 Hz	.53
Korean Mean	74.8	70.1	13.36 Hz	14.26 Hz	.9
Mandarin Mean	66	59.4	15.15 Hz	16.83 Hz	1.68
Slavic Mean	65	46.9	15.38 Hz	21.32 Hz	5.94
Somali Mean	76.6	33.1	13.05 Hz	30.21 Hz	17.16
Spanish Mean	67	46.4	14.92 Hz	21.55 Hz	6.63
GAE Mean	62	52.5	16.12 Hz	19.04 Hz	2.92

The vibration rates indicate that, overall, Japanese and Korean talkers flapped their [ɹ]s. Eight of the 10 Japanese participants and nine of the 10 Korean talkers did so. Only Japanese 8M, 10F, and Korean 11M trilled their [ɹ]s. In other words, 17 out of 20 (85%) participants produced [ɹ]s that are aurally indistinguishable from theirs [l]s. This explains why GMU transcriptionists transcribed most of their [l]s as [ɹ]s.

MARKEDNESS IN THE PRONUNCIATION OF RHOTICS

Cross-linguistically, the phonetic realizations of rhotics have been found to have both regional and social dialectalization significance. Labov's (1972, pp. 43-69) "Fourth Floor" study in three New York City department stores has helped draw attention to this phenomenon. In L2 contexts, rhotics can elicit three types of responses from hearers, two of which are sociolinguistically marked, while one is unmarked (Wardhaugh and Fuller 2015, pp. 101-102). Let's consider them briefly.

Perceptually Marked Pronunciations Caused by Confusion

Failure to distinguish clearly between laterals and rhotics has intelligibility consequences. According to Catford (1987, p. 88), the relative functional load (RFL) between [l] and [ɹ] in prevocalic positions is 83%. This high percentage can lead to serious unintelligibility. Furthermore, the existence of minimal pairs or near minimal pairs creates lexical competitions that can interfere with intelligibility. If a Japanese or Korean speaker substitutes [l] for [ɹ] or vice versa in *<bring> vs. <bling>, <fresh> vs. <flesh>, <blue>*

vs. <brew>, <frog> vs. <flog>, and <red> vs. <led>, an interlocutor may have a hard time deciphering the intended utterance. As a matter of fact, Japanese 2F, 11F, 12M, and Korean 1M pronounced <red> as <led>, while Japanese 5F and Korean 8F pronounced <frog> as <flog>.[5] It is true that the discourse context can mitigate the seriousness of the confusion, but as we all know, context is not a panacea against all misunderstandings.

Perceptually Marked Pronunciations Caused by Trilling

Trilling [r] is a double-edged sword. On the one hand, it enhances intelligibility, but on the other hand, it marks the speech as heavily accented. It enhances intelligibility because, as noted earlier, no trilled lateral has been found in any human language. Consequently, as soon as a trill is heard, English transcribers automatically perceived it as a rhotic. With regard to GAE, a trill [r] indexes the talker as having a foreign accent because in this dialect of English, the vibration rate of all rhotics is below 22 Hz. The trilling of [r]s in Arabic and Somali-accented Englishes caused GMU transcriptionists to perceive accurately all of the 128 rhotics that they produced. Yet, overall, their speech is heavily accented in part because of the strong trilled [r] that they produce. The Arabic-accented [r] is 2.44 Hz above the trilling threshold of 22 Hz. It should, however, be noted that three talkers, Arabic 1F, Arabic 35M, and Arabic 51M flapped their <r>s, while the seven remaining talkers forcefully trilled theirs. All six Somali talkers without exception trilled their [r]s vigorously, that is, 8.21 Hz above the threshold. For the speakers of these two languages, intelligibility is traded for a "heavy" or "thick" foreign accent.

Perceptually Unmarked Pronunciations

On average, the rhotics produced by Mandarin, Slavic, and Spanish talkers fall below the 22 Hz threshold. Sociolinguistically, speech items that do not attract accentedness attention are considered unmarked (Wardhaugh & Fuller, 2015, pp. 101-102). The vibration rates for these three groups of speakers may be surprising, since the rhotics in their respective native languages are classified as trills. My findings are in agreement with Lindau (1985, p. 161) who states that "Even in languages where a possible realization is a trill, not all speakers use a trill, and the speakers that do, have tap and approximant allophones as well as the trill. In the languages used in this study that were described as having an apical trill, about half of the speakers produced trills, but not for every token." She also observed that, in her study, most of the Spanish participants trilled their [r]s. My findings confirm hers. Of the 10 participants, Spanish 6M, 9M, 11F, and 13F flap their [r]s, while six others trilled theirs. The same is true for the 11 Slavic participants. Croatian 6F, Serbian 2M, Serbian 6M, and Serbian12F flapped their [r]s, while the seven others trilled theirs. The only group where the majority of the participants flapped their [r]s is Mandarin. Only two, Mandarin 6F and 12M, out of 10 trilled their [r]s.

[5] Other segmental errors occurred, such as the pronunciation of [f] as [ɸ] or [p], but we are not concerned with such errors in this paper.

SUMMARY

The analyses presented in this paper show that unintelligibility and accentedness for [l] and [ɹ] can be accounted for acoustically. The findings complement the impressionistic phonetic transcriptions made by the GMU linguists who transcribed the speech samples produced by the 67 non-native speakers who produced the lateral and rhotic segments under consideration. Reliance on well-established thresholds for the F3 of [l] and [ɹ] has allowed us to provide an acoustically sound explanation for why GAE hearers have a hard time differentiating between the [l] and [ɹ] produced by some Japanese and Korean L2 speakers of English. The vibration rate threshold for [ɹ] also explains why the [r]s that Arabic and Somali talkers produce are indexed as heavily accented even though their pronunciation does not interfere with intelligibility. Lastly, even though <r> is trilled in many native languages, some speakers tend not to transfer the full force of their native [r]s into English. As a result, their L2-accented [ɹ]s are unmarked.

ABOUT THE AUTHOR

Ettien Koffi, Ph.D. in linguistics from Indiana University, teaches linguistics at Saint Cloud State University, MN. Author of many peer-reviewed articles on various topics in linguistics and of four books: *Language Society in Biblical Times* (1996), *Paradigm Shift in Language Planning and Policy: Game Theoretic Solutions* (2012), *Applied English Syntax* (2010, 2015), and the *New Testament in Anyi Morofu* (2017), a task which took over 25 years. Specializing in acoustic phonetics, dialect variation, and emergent orthographies, his current research centers on speech acoustics of L2 English (within the Speech Intelligibility Framework), Central Minnesota English, and Anyi. He can be reached at enkoffi@stcloudstate.edu.

REFERENCES

Ash, S. (1982). *The vocalization of /L/ in Philadelphia*. Ph.D. Dissertation, University of Pennsylvania. Ann Arbor, MI: University Microfilms International.

Boersman, P., & Weenink, W. (2012). *Praat: Doing phonetics by computer*. Available online at: http://www.praat.org/.

Bradlow, Ann R., David Pisoni, Reiko Akahane-Yamada, and Yoh'ichi Tohkura. (1999). Training Japanese listeners to identify English /r/ and /l/: Perception and production. *Perception & Psychophysics* 61 (5): 977-985.

Catford, J. C. (1987). Phonetics and the teaching of pronunciation. In J. Morley (Ed.), *Current Perspectives on Pronunciation: Practices Anchored in Theory* (pp. 87-100). Washington, D.C.: Teachers of English to Speakers of Other Languages.

Espy-Wilson, C. (1997). Acoustic measures for linguistic features distinguishing the semivowels in American English. *Journal of the Acoustical Society of America* 92 (2), 736-757.

International Phonetic Association. (1999). *Handbook of the International Phonetic Association*. New York: Cambridge University Press.

Koffi, E. (2016). *Relevant acoustic phonetics of L2 English: Focus on intelligibility. Manuscript.* St. Cloud, MN.

Ladefoged, P. (1996). *Elements of Acoustic Phonetics*, 2nd edition. Chicago: The University of Chicago Press.

Ladefoged, P. & Maddieson, I. (1996). *The sounds of the world's languages*. Malden, MA: Blackwell Publishers Inc.

Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

Lee, H. B. (1999). Korean. In *Handbook of the International Phonetic Association* (pp. 120-123). New York: Cambridge University Press.

Lindau, M. (1985). The story of /r/. In V. Fromkin (Ed.), *Phonetic Linguistics: Essays in Honor of Peter Ladefoged* (pp. 157-168). New York: Academic Press.

Maddieson, I. (1984). *Patterns of sounds*. Cambridge Studies in Speech Science and Communication. New York: Cambridge University Press.

NCH Software. (2013). *WavePad Sound Editor*, Version 5.17. NCH Software Inc.: Greenwood Village, CO, USA.

Ryan, K. (2005). Grid-maker.praat. Available online at: http://www.linguistics.cla.du/faciliti/facilities/acoustic/praat.tml.

Saito, K. & Lyster, R (2012). Effects of form-focused instruction and corrective feedback on l2 pronunciation development by Japanese learners of English. *Language Learning* 62 (2):595-633.

Smith, J. G. (2010). *Acoustic properties of English /l/ and /r/ produced by Mandarin Chinese speakers.* MA Thesis: University of Toronto, Toronto, Canada.

Wardhaugh, R. & Fuller, J. 2015. *An Introduction to Sociolinguistics,* 7th edition. Malden, MA: Wiley Blackwell.

Weinberger, S. (2014). *Speech Accent Archive*. George Mason University. Retrieved from http://accent.gmu.edu.

Yoon, T-J. 2008. Stress-analysis.praat. Available online at: http://web.uvic.ca/~tyoon/resource/vq.praat.

Appendix: List of Participants

Arabic Speakers: Arabic 1F, Arabic 30F, Arabic 35M, Arabic 36M, Arabic 40M, Arabic 44F, Arabic 46M, Arabic 47M, Arabic 50M, Arabic 51M

Japanese Speakers: Japanese 2F, Japanese 3F, Japanese 4M, Japanese 5F, Japanese 8M, Japanese 9M, Japanese 10 F, Japanese 11F, Japanese 12M, Japanese 13M

Korean Speakers: Korean 1M, Korean 2F, Korean 3F, Korean 4F, Korean 8F, Korean 9M, Korean 10M, Korean 11M, Korean 12M, Korean 13F

Mandarin Speakers: Mandarin 1F, Mandarin 2F, Mandarin 3M, Mandarin 4F, Mandarin 5F, Mandarin 6F, Mandarin 8M, Mandarin 9M, Mandarin 12M, Mandarin 19M

Slavic Speakers: Croatian 1F, Croatian 2F, Croatian 4M, Croatian 5F, Croatian 6F, Serbian 1F, Serbian 2M, Serbian 6M, Serbian 11M, Serbian 12F, Serbian 14F

Somali Speakers: Somali 1F, Somali 2M, Somali 3F, Somali 4F, Somali 5M, Somali 6M

Spanish Speakers: Spanish 1M, Spanish 2M, Spanish 4F, Spanish 6M, Spanish 9M, Spanish 11F, Spanish 13F, Spanish 14F, Spanish 16M, Spanish 20M

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.

A MANDARIN SPEAKER'S INTONATIONAL EMPHASIS IN ENGLISH AND MANDARIN LECTURES

Di Liu, Boston University

Research suggests that native English speakers and non-native English speakers use different English intonation patterns (Pickering, 2004; Wennerstrom, 1998) and that prosody significantly affects English learners' intelligibility and comprehensibility (Derwing, Munro, & Wiebe, 1998). However, cross-linguistic studies of non-native English speakers' prosody production in English and in their first languages are highly limited. This study investigates the given-new stress connection (GNSC) realized in a Mandarin speaker's speech in English and in the same speaker's speech in Mandarin. The results show that the Mandarin speaker lacks the ability to effectively use pitch as a prosodic cue to mark new versus old information in English but the same speaker is able to effectively use prosodic cues to contrast new versus old information in Mandarin. Furthermore, a large portion of the new information that should be emphasized in English is emphasized in the Mandarin version but not the English version of the lecture. The findings of this study imply that Mandarin speakers may use prosodic features to emphasize new information while speaking in Mandarin but they may not fully transfer this strategy to English. The findings of this study also imply that ESL learners' first language could be a valuable resource in English prosody teaching.

INTRODUCTION

Pronunciation teaching has changed drastically in recent years (Murphy & Baker, 2015). One profound change in the contemporary field of pronunciation pedagogy is a shift in the features upon which teachers focus on in their classrooms. Pronunciation was dominated by the teaching of segmental features for many years. In the past 30 years, however, a growing number of researchers and teachers realized the significant influence that prosody has on intelligibility and comprehensibility; this led them to call for a shift of focus towards prosodic features (Anderson-Hsieh, Johnson, & Koehler, 1992; Celce-Murcia, Brinton, & Goodwin, 2010; Derwing, Munro, & Wiebe, 1998).

PROSODY

Often described as the "music" or "melody" of language (Allen, 1971; Wennerstrom, 2001), prosody encompasses a range of linguistic variables including intonation, stress, pauses, and rhythm. Researchers generally agree that prosody encodes meanings and has various pragmatic functions (Levis & Wichmann, 2015; Pierrehumbert & Hirschberg, 1990; Reed & Michaud, 2005; Wennerstrom, 2001). At the lexical level, prosody can change the categories of lexical items, for example, *REcord* with lexical stress on the first syllable is a noun and refers to something permanent and *reCORD* with lexical stress on the second syllable is a verb and means to convert something into permanent form.

At the phrasal level, prosody has much broader functions and a less fixed influence on meaning. A sophisticated understanding of prosodic elements is important because they mark information structure, including the distinction between new and old constituents (Hahn, 2004; Pickering, 2004; Pierrehumbert & Hirschberg, 1990; Wennerstrom, 2001). At the phrasal level, prosody shifts the focus of utterances and can mark contrastive stress. Stress can change not only the meaning of a sentence, but also its focus and implication. For example: *It is not MY responsibility to do this job* and *It is not my RESPONSIBILITY to do this job* realize focus on different constituents, and the implications differ due to stress alone. The former sentence suggests that it is someone else's responsibility to do the job, whereas the latter emphasizes the fact that the speaker is not obliged to do the job. Hahn (2004) demonstrated that sentence level primary stress plays an important role in intelligibility by investigating native speakers' reaction times, memory and attitudes to sentences read by an international teaching assistant "with primary stress correctly placed (e.g., *The URban environment is more individualistic than the RUral environment*.), misplaced (e.g., *The urban environment is more individualistic than the rural enVIronment*.) or missing (e.g., *The urban environment is more individualistic than the rural environment*)" (p. 206). She found that when the primary stress was correctly placed, native listeners processed the lecture more easily, recalled significantly more information, and provided significantly better speaker evaluations.

PROSODY FUNCTIONS IN CLASSROOM

Researchers have noted the pragmatic functions of prosody in classroom. Chun (1988) states that, in classrooms, teachers use a wide range of communicative functions including, "addressing others, selecting the next speaker, choosing the topic, interrupting, asking for clarification, changing the subject, and concluding a discussion" (p. 82). 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 and that teachers use complex prosody packaging while providing feedback to students. Wennerstrom (1998) proposed that there is an intonation system in English that functions at the discourse level to signal relationships in information structure. She proposes a model in which intonation functions as a grammar of cohesion.

ITAS' CLASSROOM PROSODY

Pickering compared native speaker teaching assistants' (NSTAs') and international teaching assistants' (ITAs') tone choice and use of intonation in the classroom. She argues that, "intonation bears a high communicative load in terms of information structuring and rapport building between discourse participants" (2001, p. 234). Pickering's (2004) subsequent 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 are generally incapable of effectively controlling this level of structural organization. Specifically, ITAs in her studies made use of an overall narrower pitch range and were unable to consistently manipulate *key*, which is defined by Brazil (1981) as the starting pitch level of an intonation unit, and *tone choices* to create intonational paragraphs.

PROSODY TEACHING

Prosody is hard to teach due to its fluid nature and dependence on contextual information. Previous literature that intends to generalize the "meanings" of pitch contours concludes that the "meanings" are highly context-dependent and that even native speakers can hardly reach a consensus of the "meanings" of pitch contours (Lieberman, 1967, p. 124). Native English speakers gain this contextual information and knowledge over many years and through frequent communication with other native speakers. However, it is challenging for non-native speakers of English to gather this contextual information in a much shorter learning period with much less communication with native speakers of English.

One possible solution to the issue regarding contextual information is to transfer certain prosodic functions from one language to another. If we consider the situation for Mandarin learners of English, there is a relative lack of empirical studies that directly compare Mandarin speakers' English and Mandarin use of similar functions across their two languages. It is not until recently that researchers have started to investigate cross-linguistic functions of prosody between Mandarin and English.

MANDARIN PROSODY

Ouyang and Kaiser (2015) conducted a study to investigate corrective focus and given/new contrast in Mandarin prosody. The results show that even a language with lexical tones (Mandarin), uses prosodic features (i.e., fundamental frequency (F0), duration and intensity) that are used to encode information structure in English. In fact, not only can prosodic cues indicate importance at the discourse level, they also distinguish different types of information structure in Mandarin. Analyzing telephone dialogue data, Ward, Li, Zhao and Kawahara (2016) conclude "a significant fraction of the prosody of Mandarin dialog relates to pragmatic functions", which include, but are not limited to "back-channeling, turn yielding, turn starts, negative evaluation, bids for empathy, topic progression, topic exhaustion, and contrast" (p. 1234). Of course, it is still possible that even though Mandarin and English both use the same suprasegmental features in realizing the same functions, they do so to different degrees. However, a recent study presented by Ip and Cutler (2016) suggests that pitch levels used to mark focus in Mandarin are even higher than they are English when marking focus. Comparing five different types of prosodic focus by native English speakers and Mandarin speakers, Ip and Cutler find that "native speakers of Mandarin resemble English speakers in their tendency to signal focus by manipulation of duration, pitch range, and intensity." (p. 333) Furthermore, they find that "Mandarin speakers reliably show greater increase in pitch for new-information focus (either as pitch range or mean/maximum)" (p. 333). Although these studies suggest that there are similarities between Mandarin and English prosody, studies that directly compare the same speaker's prosody production in English and Mandarin are lacking.

GNSC EFFECT

Among all the prosodic features, stress may be the most important and widely-studied one. Stress constituents usually have a higher pitch level, longer duration, greater amplitude and full vowels. One particular function of stress that warrants attention is the "GNSC" effect. Pierrehumbert and Hirschberg (1990) argue that in English, new and contrastive information tends to be realized with a high pitch accent, and old or given information tends to be realized with a low pitch accent. Hahn (2004) refers to this relationship as the given-new stress connection (GNSC). The present study investigates if the GNSC is demonstrated in a Mandarin speaker's speech in English and if the same relationship is shown in the speaker's speech in Mandarin. The research questions are:

1. How does a speaker who speaks both Mandarin and English demonstrate the given-new stress connection (GNSC) in Mandarin and English?
2. To what degree is the given-new stress connection (GNSC) realized in a native Mandarin speaker's Mandarin lecture?
3. To what degree is the given-new stress connection (GNSC) realized in a native Mandarin speaker's English lecture?

METHOD

The material analyzed is a massive open online course (MOOC): *Principle of Electric Circuits* retrieved from EdX. This course is delivered in two languages (Mandarin and English) by the same speaker, who is a native speaker of Mandarin and who was a visiting scholar at the Massachusetts Institute of Technology. The English version is a direct translation of the Mandarin version, and the speaker made little modification of the content while he was giving the lecture using PowerPoint slides. The recordings were made for the same purpose and the same online platform, so that the qualities of the sound files are similar.

The lectures were downloaded from EdX and analyzed and annotated using the speech analysis software *Praat*. The speech samples are annotated based on the ToBI (Tone and Break Indices) conventions (Beckman & Hirschberg, 1994). The English version of the lecture was annotated based on the MAE (Mainstream American English)_ToBI conventions (see Figure 1).

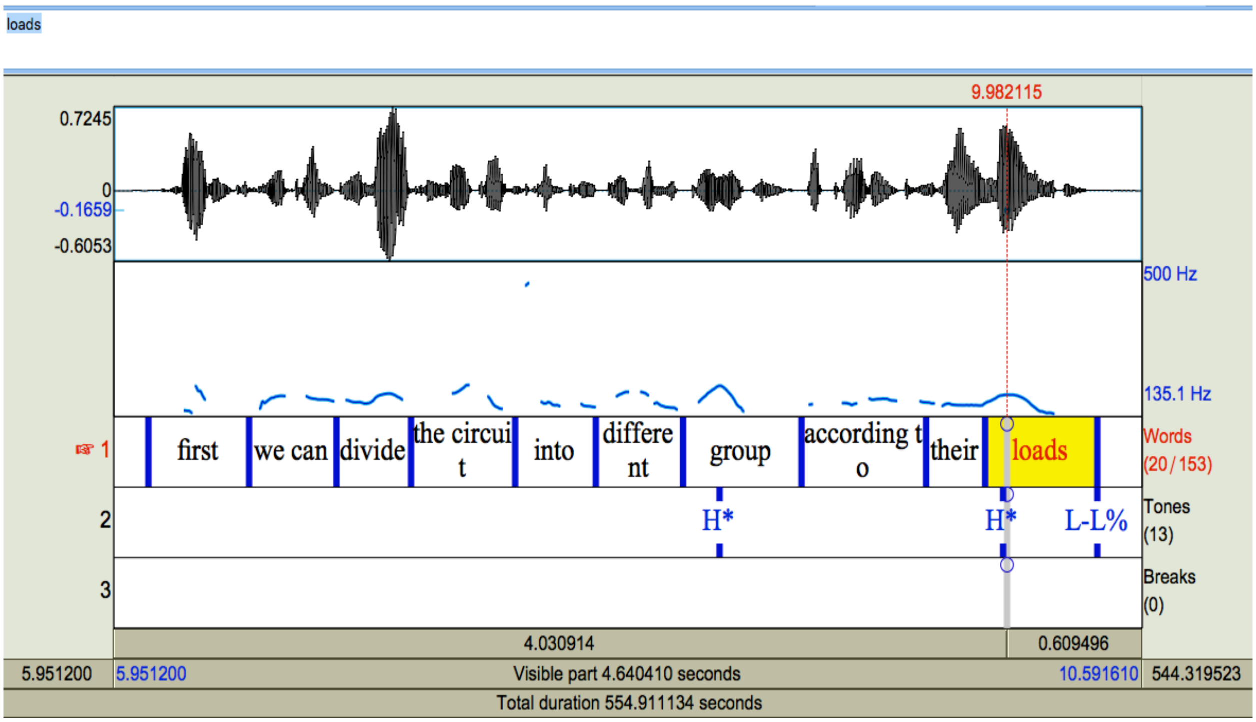


Figure 1. Sample MAE_ToBI Transcription.

The Mandarin version of the lecture was annotated according to the Pan Mandarin_ToBI conventions (see Figure 2).

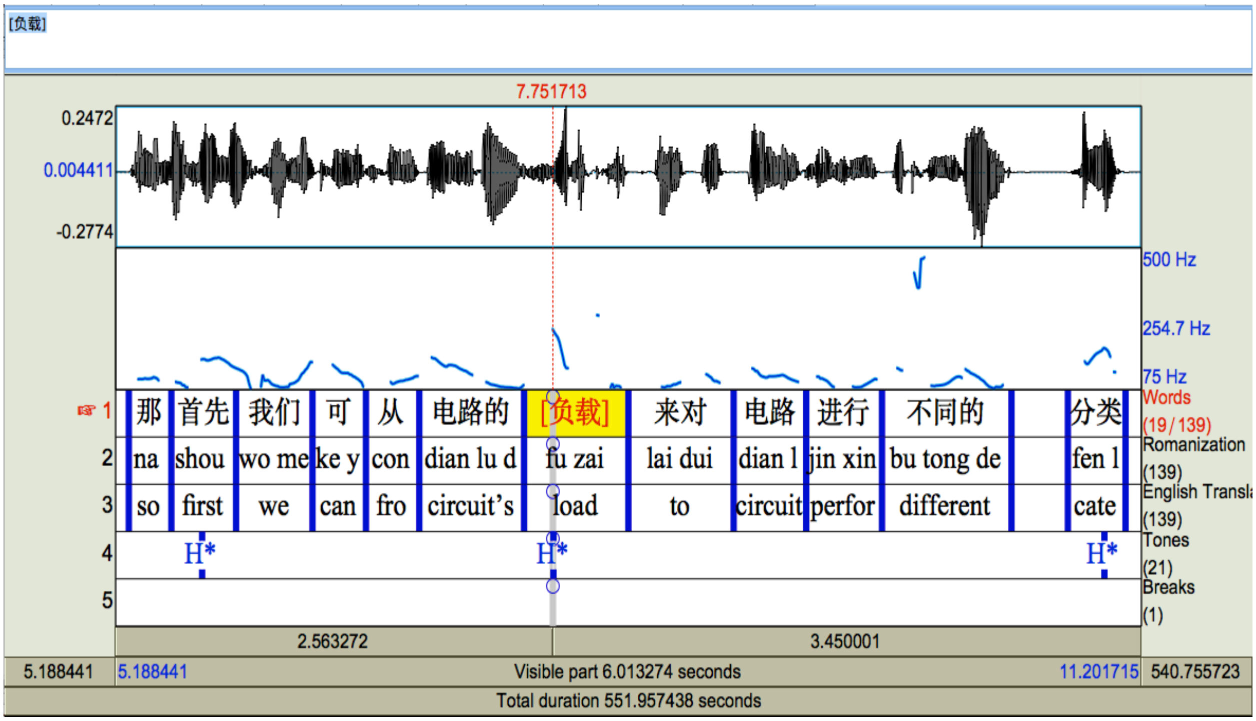


Figure 2. Sample Pan Mandarin_ToBI Transcription.

Because the goal of this research is to analyze items that were realized as both "new" and given," 32 concepts that were mentioned more than one time are analyzed below. If the concept is mentioned the first time in the lecture, it was coded as "new" and if the concept has been mentioned before in the lecture, it was coded as "old" (e.g. load $_{new}$ vs. load $_{old}$). The maximum pitch level of these words and phrases were retrieved using the built-in pitch elicitation function in *Praat*. Paired t-tests were conducted to compare the maximum pitch level across different groups.

RESULTS

A paired t-test was conducted to determine if there was a difference in the mean maximum pitch level of the Mandarin words/phrases as compared to the mean maximum pitch level of the English words/phrases. The total number of words/phrases analyzed is 64 for each group (32 new concepts and 32 old concepts). The Mandarin version of the lecture has a higher mean pitch level (M = 176.25) as well as a higher standard deviation (SD = 53.49) compared to the mean pitch level (M = 160.99) and the standard deviation (SD = 30.42) of the English version of the lecture, suggesting that the Mandarin speaker uses a higher pitch level and that his pitch fluctuates more when he speaks in Mandarin than when he speaks in English. A paired t-test was conducted to determine if this difference was statistically significant (t = 2.0362, df = 63, p = 0.04594). The null hypothesis is rejected, and we can conclude that there is a statistically significant difference between the pitch level of words/phrases in the speaker's Mandarin speech and English speech and the speaker's Mandarin speech has a significantly higher pitch level compared to the same speaker's English speech.

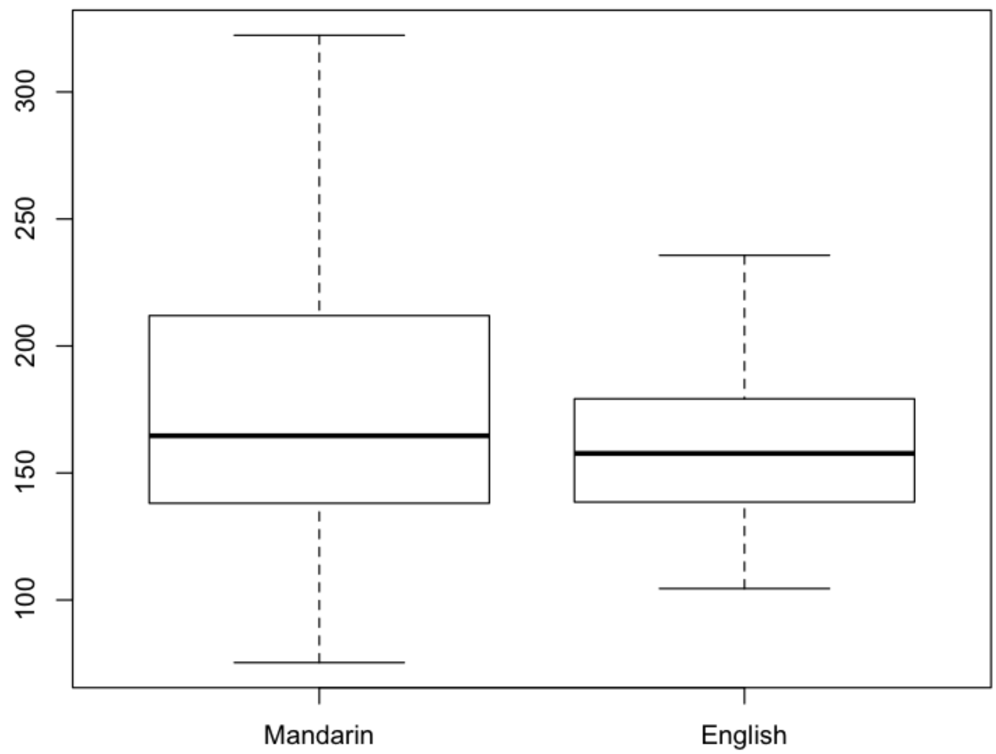


Figure 3. The speaker's mean maximum pitch levels (in Hz) in Mandarin and English.

A paired t-test was conducted to determine if there was a difference in the mean maximum pitch level of the new concepts as compared to the mean maximum pitch level of the old concepts in the Mandarin lecture. The total number of words/phrases analyzed is 32 for each type of information. The mean and standard deviation of new information (M = 195.09, SD = 47.32) are significantly different from those of the old information (M = 157.42, SD = 47.32). A paired t-test determined that this difference was statistically significant (t = 3.0707, df = 31, p = 0.0044). The null hypothesis is rejected, and we can conclude that there is a statistically significant difference between the pitch level of the new information and old information in the speaker's Mandarin speech. The pitch level of the new information in the speaker's Mandarin speech is significantly higher than the pitch level of the old information in the same speaker's Mandarin speech.

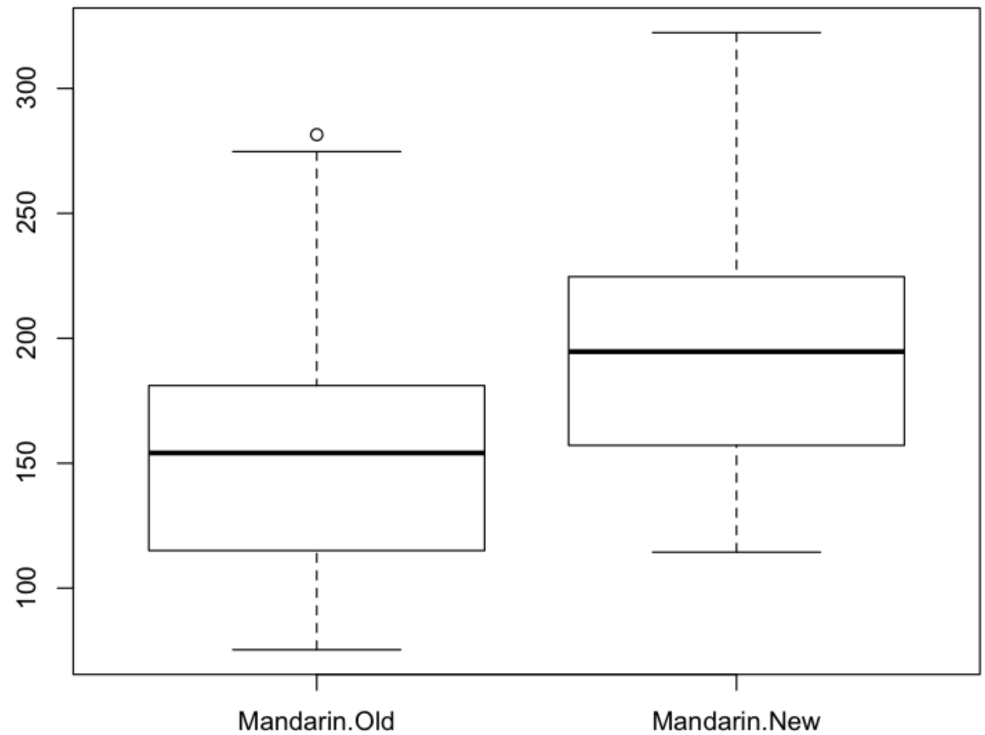


Figure 4. The mean maximum pitch levels (in Hz) of Mandarin old and new information

A paired t-test was also conducted to determine if there was a difference in the mean maximum pitch level of the new concepts as compared to the mean maximum pitch level of the old concepts in the English lecture. The total number of words/phrases analyzed is 32 for each type. The mean and standard deviation of new information (M = 164.83, SD = 28.16) are close to those of the old information (M = 157.16, SD = 32.52). A paired t-test was conducted to determine if this difference was statistically significant, and it was not ($t = 1.4744$, $df = 31$, $p = 0.1505$). This lack of statistically significant difference between the pitch levels of new and old information in the English lecture suggests that GNSC effect in the Mandarin speaker’s English speech is not significant.

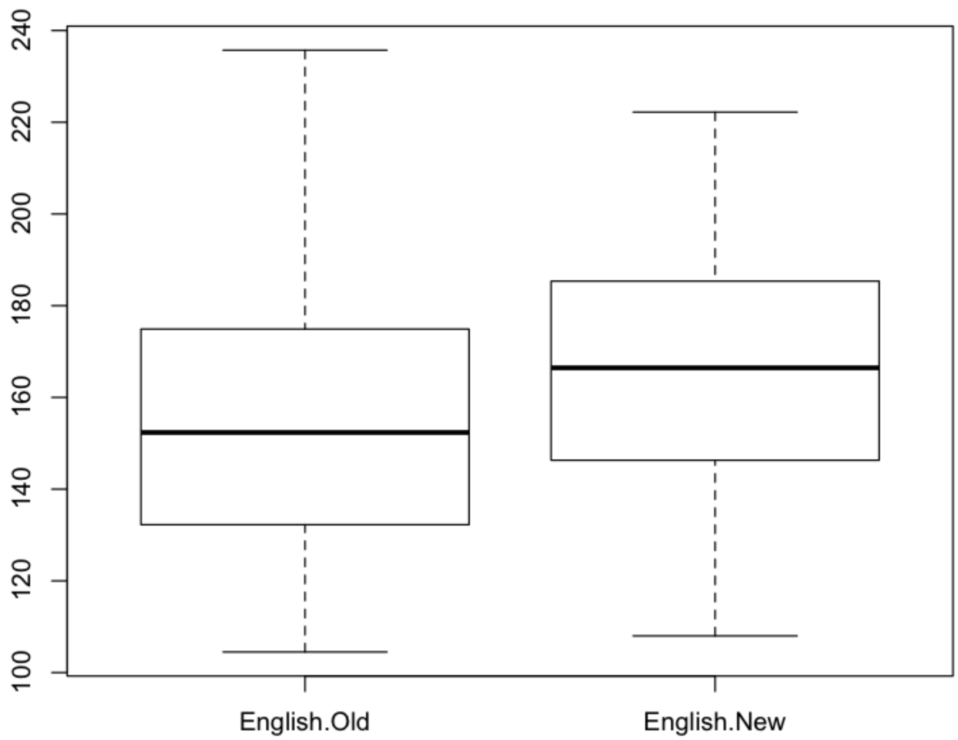


Figure 5. The mean maximum pitch levels (in Hz) of English old and new information

CONCLUSION

This study investigates the given-new stress connection (GNSC) realized in two versions (English and Mandarin) of the same lecture given by a native Mandarin speaker who is fluent in both English and Mandarin. The results show that the Mandarin speaker uses a higher pitch level and a more varied pitch contour while delivering the lecture in Mandarin. In terms of the GNSC, the Mandarin speaker demonstrates significantly different pitch levels between new and old information in the Mandarin version of the lecture but does not demonstrate statistically significantly different pitch levels when he was delivering the same lecture in English. This result suggests that the Mandarin speaker lacks the ability to effectively use prosodic cues to contrast new versus old information in his English lecture even though the speaker is able to use prosodic cues to contrast new versus old information effectively when giving the same content in Mandarin. In other words, the Mandarin speaker does not fully transfer his skills and strategies of using prosodic cues from Mandarin to English.

Students' lack of awareness of the importance of prosody hinders prosody learning. As Gilbert (2014) said, "[w]hile it might seem self-evident to a teacher that intonation is important, students may simply consider imitating pitch contours 'awkward,' even if they seldom tell teachers that they 'feel silly' speaking in this way. As a result of these differing perspectives, students "may walk out of the class without having accepted the system at all" (p.125). Gilbert's observation echoes what Allen (1971) noted 40 years ago: "there is little carry-over into the students' own conversations outside the classroom and the listen and repeat approach has never yielded satisfactory long-term results" (p. 79). The fact that this issue remains unsolved after so many years even though researchers have acknowledged it suggests that we may lack effective techniques to enhance learners' awareness of the importance of English prosody. The findings of this study, however, suggest a possible route. If students understand that they have been using prosody to realize a range of pragmatic functions in their first language(s) and that English uses prosody for the same purposes, it may be much easier for them to accept the system and adopt English prosody.

This study also finds that even two languages (Mandarin and English) that differ significantly in lexical level prosody (i.e., Mandarin is a tone language and English is a non-tone language) share similar prosodic functions. Therefore, based on the findings of this study, English learners' L1s might be a valuable resource in prosody teaching. Pronunciation teachers might be able to improve the intelligibility and comprehensibility of the learners by facilitating positive prosodic transfer between students' L1s and English discourse prosody functions. Future studies should investigate prosodic functions in other languages and compare the relatively well-studied English prosody functions with the less-studied prosody functions of other languages. Finally, the findings of this study are based on the analysis of a single subject's speech therefore might lack generalizability. Future studies should include more participants to confirm or contradict the findings of this study.

ABOUT THE AUTHOR

Di Liu is a third-year doctoral student in Boston University. He is working with Dr. Marnie Reed on prosody research and teaching. His research includes (1) cross-linguistic comparison and transfer of Mandarin and English classroom prosody; (2) asymmetrical cognitive load imposed by processing native and nonnative speech and (3) constraints on downtrend as a function of tone inventory structure: evidence from Cantonese and Mandarin. Serving as the TA of a Computer-Assisted Language Learning course in BU for the past three years, he is also designing a curriculum/textbook: Technology-Enhanced Pronunciation Teaching, which intends to provide a systematic pedagogical application of various software/websites/apps that are available to teachers.

520 Main Street, Apt. 901
Malden, MA, 02148
917-868-5796
diliu@bu.edu

REFERENCES

Allen, V. (1971). Teaching intonation, from theory to practice. *TESOL Quarterly, 5*(1), 73-81.

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.

Beckman, M. E, & Hirschberg, J. (1994). The ToBI annotation conventions. Online MS. Available at http://www.ling.ohio-state.edu/tobi/ame.tobi/annotation.conventions.html

Brazil, D. (1981). The place of intonation in a discourse model. In R. M. Coulthard & M., Montgomery (Eds.), *Studies in Discourse Analysis*. (pp. 146-157). London: Routledge & Kegan Paul.

Celce-Murcia, M., Brinton, D., & Goodwin, J. M. (2010). *Teaching pronunciation: A course book and reference guide*. New York: Cambridge University Press.

Chun, D. (1988). Teaching intonation as part of communicative competence: Suggestions for the classroom. *Die Unterrichtspraxis / Teaching German, 21*(1), 81-88.

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

Gilbert, J. (2014). Myth 4: Intonation is hard to teach. In L. Grant, (Ed.), *Pronunciation myths: Applying second language research to classroom teaching*. Ann Arbor, MI: University of Michigan Press.

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

Ip, M. H. K., & Cutler, A. (2016). Cross-language data on five types of prosodic focus. *Proceedings of Speech Prosody 2016*, May 31-Jun 3, Boston

Hellermann, J. (2003). The interactive work of prosody in the IRF exchange: Teacher repetition in feedback moves. *Language in Society, 32*(1), 79-104.

Levis, J. M. & Wichmann, A. (2015). English intonation: Form and meaning. In M. Reed & J. Levis (Eds.). *The Handbook of English Pronunciation* (pp. 139-155), West Sussex, UK: John Wiley & Sons, Inc.

Lieberman, P. (1967). *Intonation, perception, and language* (M.I.T. research monograph 38). Cambridge: M.I.T. Press.

Murphy, J. & Baker, A. (2015). History of ESL pronunciation teaching. In M. Reed & J. Levis (Eds.). *The Handbook of English Pronunciation* (pp. 36-65), West Sussex, UK: John Wiley & Sons, Inc.

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*, 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.

Pierrehumbert, J. & Hirschberg J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. R., Cohen, J. Morgan, & M. E. Pollack (Eds.) *Intentions in Communication* (pp. 271-323). Cambridge, Massachusetts: The MIT Press.

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.

Ward, N. G., Li, Y., Zhao, T., & Kawahara, T. (2016). Interactional and pragmatics-related prosodic patterns in Mandarin dialog. *Proceedings of Speech Prosody 2016*, May 31-Jun 3, Boston.

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; New York: Oxford University Press.

Piccardo, E. & North, B. (2017). Developing phonology descriptors for the Common European Framework of Reference (CEFR). 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. 97-109). Ames, IA: Iowa State University.

DEVELOPING PHONOLOGY DESCRIPTORS FOR THE COMMON EUROPEAN FRAMEWORK OF REFERENCE (CEFR)

Enrica Piccardo, OISE-University of Toronto, Université Grenoble-Alpes – France
Brian North, Eurocentres, Eaquals

While pronunciation is the object of an increasing number of studies, the translation of this growing body of research into teacher-oriented resources is still at its initial stages (Derwing & Munro, 2015). Often still playing *de facto* a gatekeeping role in society, the importance of phonological competence is being more and more recognized, with explicit instruction increasingly appearing in the L2 classroom, and a focus on the definition of valid assessment criteria for pronunciation proficiency. However, there is still a lack of validated, concrete descriptors to support practitioners and learners alike. Although the Common European Framework of Reference for Languages (CEFR, 2001) presents a solid conceptualization of phonological competence, this did not aptly translate into the existing scale for *Phonological control*. In the context of a Council of Europe's project to update the CEFR descriptors, a new analytic scale was developed to replace the existing holistic one. This article reports on the methodology used for developing and validating the descriptors for *Overall phonological control*, *Sound articulation*, and *Prosodic features*. The rationale behind the scale is discussed together with the complex relationship between language proficiency and phonological competence, in the light of the notion of intelligibility.

CONTEXT

The *Common European Framework of Reference for Languages: Learning, teaching, assessment* (CEFR) (Council of Europe, 2001) is an influential language policy document in Europe and beyond, and its descriptor scales, empirically developed in a Swiss National Research Project (North & Schneider 1998; North 2000; Schneider & North 2000), are used as a source for the development of language standards. Among the wealth of scales offered by the CEFR, only one scale deals with phonological competence, the phonological control scale (Council of Europe, 2001, p. 117). Since the publication of the CEFR in 2001, unlike the other scales included in the document, the scale for phonological control has attracted a great deal the criticism, particularly from researchers interested in both phonology and the CEFR (see in particular Horner, 2010, 2013, 2014, Frost & O'Donnell, in press, Harding, 2016). This does not come as a real surprise, though, as problems had been encountered with the scaling of descriptors for phonology in the Swiss research project previously mentioned, partly because it "can involve an implicit negative concept: that of accent. Less accent is good, more accent is bad" (North 2000, p. 238). Furthermore, in the Rasch analysis to calibrate the descriptors, phonology did not 'fit' well in the construct of the rest of the descriptors for spoken English, and then 'misfitted' completely when other languages (French and German) were added.

The need for the creation of a new illustrative descriptor scale for phonological control to be integrated in the CEFR appeared both timely and critical. The opportunity to address this issue was offered by a 2013–2016 Council of Europe project aimed at extending the CEFR illustrative descriptors. The focus of the overall project was to provide scales for some areas that had not been covered in the 2001 publication, particularly mediation. The conceptualization of mediation addressed aspects of language teaching pedagogy that had been present but underdeveloped in the CEFR (Piccardo, 2012) and took into consideration new, related perspectives that have been emerging in language education.

One underdeveloped area was precisely phonology. Although the CEFR presents a solid conceptualization of phonological competence, albeit necessarily adapted to an audience of language educators and not of phoneticians, this conceptualization was not adequately operationalized in the existing scale. Needless to say the existing scale did not incorporate the new vision of phonology that has developed since the mid-1990s when the CEFR research was undertaken, as will be discussed below. The project reported on aimed to address some of these issues and to provide realistic, concrete descriptor scales to support practitioners and learners alike. Work was undertaken in four phases: analysis of existing instruments, literature review, development and initial feedback, qualitative and quantitative validation.

ANALYSIS OF EXISTING INSTRUMENTS

Naturally the starting point was an analysis of the treatment of phonology in the CEFR. Phonological competence has an important place in the descriptive scheme of the CEFR. It is mentioned in the overview of **communicative competences** (Section 2.1.2, p. 22) and in the description of the **communicative language processes** (Section 4.5, pp. 90–91), when it comes to execution (Section 4.5.2) and especially Production (4.5.2.1):

> "The production process involves two components: The *formulation* component takes the output from the planning component and assembles it into linguistic form. This involves lexical, grammatical, phonological (and in the case of writing, orthographic) processes which are distinguishable and appear (e.g. in cases of dysphasia) to have some degree of independence but whose exact interrelation is not fully understood. The *articulation* component organises the motor innervation of the vocal apparatus to convert the output of the phonological processes into coordinated movements of the speech organs to produce a train of speech waves constituting the spoken utterance, or alternatively the motor innervation of the musculature of the hand to produce hand- written or typewritten text."
>
> (Council of Europe 2001, p. 91)

The main coverage then comes in the description of **communicative language competences** (Section 5.2), where it is detailed as follows:

> "5.2.1.4 Phonological competence involves a knowledge of, and skill in the perception and production of:

- the sound-units (*phonemes*) of the language and their realisation in particular contexts (*allophones*);
- the phonetic features which distinguish phonemes (*distinctive feature*s, e.g. voicing, rounding, nasality, plosion);
- the phonetic composition of words (*syllable structure*, the sequence of phonemes, word stress, word tones);
- sentence phonetics (*prosody*):
    - sentence stress and rhythm;
    - intonation;
- phonetic reduction:
    - vowel reduction;
    - strong and weak forms;
    - assimilation;
    - elision. (Council of Europe 2001, p. 116-117)

In addition, phonological control is listed among the 12 qualitative categories relevant to oral assessment (Section 9.4, p. 193). Finally, there is the descriptor scale already mentioned, which provides one descriptor per level from A1 to C1 (p.117). The strengths and weaknesses of the treatment of phonology in the CEFR can thus be summarized as follows:

Strengths

The description of phonology in the CEFR is thorough and sufficiently broad to allow a revision and extension of the descriptors in order to capture recent reflections and developments in second/foreign language education. Learnability (and consequently 'teachability') of phonology is envisaged. The integration of phonology among relevant categories of assessment is clearly mentioned, especially in the appendices.

Weaknesses

The existing scale *Phonological Control* does not capture this conceptual apparatus, and appears unrealistic with regard to accent and progression (particularly in an apparent leap between B1 *Pronunciation is clearly intelligible even if a foreign accent is sometimes evident and occasional mispronunciations occur* (actually calibrated to B1+: North 2000) and B2 *Has a clear, natural, pronunciation and intonation.* (actually calibrated to B2+: North 2000). This latter descriptor also suggests a native speaker standard as the criterion for 'natural,' the only CEFR descriptor to compare to the competence of a 'native speaker.' Finally, the scale mixes stress/intonation, pronunciation, accent and intelligibility without providing clear indication of progression in any of these factors specifically.

A number of studies have been published on the CEFR phonology scale (Cauvin, 2012; Galaczi, Post, Li, & Graham, 2011; Horner, 2010, 2013, 2014; Isaacs & Trofimovich, 2012; Harding, 2013). Horner examines which pronunciation features have an impact on communicative efficiency, questioning the possibility of establishing hierarchies and analyzing the aspects of pronunciation that are "irritating" (and for whom), difficult to acquire (and in what order), and that have an impact on intelligibility. With regard to intelligibility, he identifies as core features, factors such as word stress and accurate reproduction of phonemes and sentence stress, while he considers intonation, rhythm, and phonetic reduction as more peripheral. Although he refers to English language phonology in his articles, these considerations can apply across languages, albeit with individual variations. Another important study is Trofimovich and Isaacs, in which the authors underline how the CEFR combination of "descriptions of easily understandable speech and a noticeable foreign accent in the same band descriptor" (2012, p. 914) can be problematic and stress the need to "disentangle accent from different aspects of communicative effectiveness, including comprehensibility."(2012, p. 914). Finally, Harding (2016) criticises the usability of the current scale. He used mixed-methods research to investigate the construct underlying the scale, particularly its orientation towards a nativeness principle or a comprehensibility principle respectively.

Other language proficiency scales, however, are not noticeably better than the CEFR scale in their treatment of phonology. In assessment, phonology is rarely a feature to be separately assessed, being generally subsumed under "speaking" or "fluency". In the TOEFL iBT Test, for instance, references to pronunciation are to be found under the category "Delivery", which includes in the same descriptor four different, albeit overlapping, features (fluency, pronunciation, intonation, and intelligibility). In the Hong Kong Diploma of Secondary Education (HKDE), specific descriptors are provided organized around 5 levels of "pronunciation and delivery." Only the IELTS test defines 9 bands of "pronunciation." Then there is the issue of the way that progression is achieved in scales for phonology through the substitution of qualifiers like "some", "a few", "many", "most", etc. The inadequacy of such an approach has long been acknowledged (Alderson 1991; Champney 1941,). Consistent interpretation becomes impossible unless raters are trained to interpret the descriptors in the same way, which they will then tend to do without referring to the actually wording. This issue is particularly acute in the IELTS scale (e.g., "full range of pronunciation features", "wide range", "range", "limited range", "some features"). Unfortunately, the scales produced by researchers trying to improve on the CEFR phonology scale (e.g. Frost & O'Donnell, in press; Horner 2013) tend to continue in the same vein.

LITERATURE REVIEW

In addition to studying papers concerning the CEFR phonology scale, a broader literature review was conducted (using www.teacherreference.com, www.mla.org; ERIC, and Google scholar). The aim was to explore the way phonology is dealt with, from a pedagogical point of view in second/foreign language education; and to identify resources and concepts to inform the development of descriptors.

Strikingly few of the articles found focused specifically on the teaching, learning, and/or assessment of phonology. The research tradition in phonology has often been only peripherally relevant to applied linguistics (Derwing & Munro, 2015). Indeed, as Munro & Derwing (2011) point out in their research timeline on accent and intelligibility, there is a fundamental incongruence between research interests and pedagogical ones. The focus in research on accent and accuracy instead of on intelligibility has been detrimental to the development of pedagogically-oriented work on pronunciation. "Identifying nativelike production as the central goal in pronunciation teaching inevitably leads to the conclusion that pronunciation is probably not worth teaching because of the limited likelihood of achieving that end." (Munro & Derwing, 2011. p. 317). Even though intelligibility has been seen as a priority in language instruction for a long time (the authors mention Sweet's handbook dating back to 1900!), conceptualization of the notion is a relatively recent process, even though "it is intelligibility – rather than native-like pronunciation – that is most critical for successful communication in an L2" (Munro & Derwing, 2011, p. 316–317).

Although there has been a growing body of such research in the last few years (e.g. Saito, 2012; Lee, Jang & Plonsky, 2015; Thomson & Derwing, 2014), research and reflection on the role of phonology in second/foreign language education is still limited, and so are resources to support teachers. In a seminal article, Derwing and Munro described the situation as follows: "The study of pronunciation has been marginalized within the field of applied linguistics [and] as a result, teachers are often left to rely on their own intuitions with little direction" (2005: 379). Derwing and Munro (2015) report, on the basis of a wide range of studies, that "teachers are hesitant about systematically teaching pronunciation" (p. 78), that they feel a "need for access to more professional development" (p. 80) and that "[t]he curricula in the various programs in which the teachers worked did not focus on pronunciation" (ibid.). Teachers do not teach pronunciation because they "lack confidence, skills and knowledge" (MacDonald, 2002: 3). They have received little or no specific training (Breitkreutz, Derwing & Rossiter, 2002; Burgess & Spencer, 2000; Derwing & Munro, 2005; Baker, 2011).

Derwing & Munro list a series of pronunciation-related phenomena that have been identified by various researchers as crucial to the success of any communicative activity. These include "listener factors, such as familiarity with foreign-accented speech, willingness to communicate, and attitudes toward L2 speakers" (Derwing & Munro, 2005, p. 392). However, "little research has explored which linguistic features of speech are most crucial for intelligibility and which, while noticeable or irritating, merely contribute to the perception of an accent" (Trofimovich & Isaacs, 2012, p. 906). This is unfortunate "given the well-established fact that accent is partially independent of comprehensibility and intelligibility and that the latter two are more important to successful communication" (Derwing & Munro 2015, p. 168). Some studies are starting to focus on assessment, in order to address the fact that proficient speakers can be unfairly penalised due to the mismatch between their level of functional proficiency and their level of phonological competence. Isaacs (2008) and Trofimovich & Isaacs (2012) highlight the centrality of intelligibility and attempt to provide rating scale specifications that move beyond the native speaker standard and bring light to what are often just intuitive impressions (Isaacs & Thomson, 2013).

DEVELOPMENT AND INITIAL FEEDBACK

The literature review helped to identify categories that could inform scales based on developments in pronunciation research and teaching. It was also insightful in thinking through the internal progression of phonological competence as well as its relationship to progression in language proficiency. In drafting descriptors, the aims were (a) to revise the general scale for phonological control to provide a snapshot of phonological competence; and (b) to supplement this with sub-scales to help teachers and learners to identify areas for improvement. The following core areas to inform descriptor production were identified:

- **Articulation** (including pronunciation of sounds/phonemes);
- **Prosody** (including intonation, rhythm and stress –word stress/sentence stress – and speech rate/chunking);
- **Accentedness** (accent and deviation from a ‘norm’);
- **Intelligibility** (i.e., actual understanding of an utterance by a listener) and **comprehensibility** (i.e., listener’s perceived difficulty in understanding an utterance).

It proved difficult to establish a clear distinction between these aspects, since they overlap and influence judgment in an inextricable way. It was therefore decided to operationalize them in only three scales: *Overall phonological control*; *Sound recognition and articulation*; *Prosody (intonation, stress and rhythm)*. Intelligibility was the key factor for discriminating between levels and explicit mention of accentedness was made at all levels in the *Overall* scale in order to raise awareness of the non-causal relationship between accent and phonological competence.

First drafts were piloted in a one-hour workshop with eleven English and French language experts from different European countries, in order to check the clarity and appropriateness of the descriptors, the coherence of the specific aspects they intended to assess and the coverage of relevant issues. The participants were asked in pairs to work on two worksheets. One contained mixed descriptors from both the prosody/intonation scale and the sound articulation scale, presented in random order; the other contained the descriptors for overall phonological competence, again in random order. Participants were asked to identify the descriptors belonging to each of the scales, assign the descriptors of each identified scale to the six CEFR levels, and provide qualitative feedback on the scales and/or the individual descriptors. The results were encouraging: only one or two sound articulation / prosody descriptors were misplaced per worksheet, which appeared to confirm the clarity and consistency of the scales. The overall descriptors were correctly matched to the intended CEFR levels with no disagreement. The final task, ranking the sub-scale descriptors by level, generated more discussion, but produced almost 70% of correct ranking. After completion of the tasks, an extensive discussion followed in which participants provided qualitative feedback on both the descriptors and the scales. These insights were very helpful in revising them.

The revised scales were then shared with five other experts working on phonology in relation to the English and French languages. Reactions were very positive, some consultants giving very detailed feedback. The specific feedback on each descriptor that was received was collated and compared in order to inform revision. On the basis of this feedback, and through wider consultation within the authoring group of the main descriptor extension project, the descriptors were again revised, translated into French, and prepared for the validation process.

QUALITATIVE AND QUANTITATIVE VALIDATION

The validation took place in two phases between January 2, 2016 and February 7, 2016. In a first—qualitative—phase, the descriptors were rated for quality before being matched the intended categories and levels. In the second—quantitative—phase, data from judgements of the CEFR level of the descriptors, plus from the use of the descriptors as a checklist to assess performance on videos, was used to calibrate the descriptors to the CEFR levels. Some 250 respondents completed the survey in Phase 1. They were asked to: (a) identify the category of the descriptors; (b) rate the descriptors for clarity and for pedagogical usefulness, and (c) assign the descriptors to CEFR levels. The first task had a tricky element, in that descriptors were included (e.g. on accent) that did not fit under either of the two categories given: 'Sound articulation' and 'Prosodic features.' They were intended to land in a "Can't decide" column. Only one descriptor was not assigned to the correct category, though another three were not as clearly identified with their category as was desirable. Only two descriptors were rated at less than 80% for the two quality criteria: clarity and usefulness. In terms of assignment to level, only three descriptors were assigned to a wide range of level. As a result of the exercise, five descriptors were dropped and several others reformulated in preparation for Phase 2.

In the quantitative phase, 272 respondents took part in the survey. The first task was to assign each of 34 descriptors to a CEFR level, by answering the following question: *At what CEFR level do you think a person can do what is defined in the descriptor?* The second task consisted of assessing video performances with a rating scale applied to each descriptor.

Assigning Levels

The data was subjected to two analyses: 1) a collation showing, for each descriptor, the percentage that rated the intended level, and 2) a Rasch analysis (Linacre, 2014) to link the ratings to the mathematical (logit) scale underlying the CEFR levels (North 2000). In the simple collation, 28 descriptors were assigned to the intended level by 40% or more of the respondents. Five descriptors seemed to be relatively evenly divided between two levels, e.g. *Can articulate the majority of the sounds of the target language reasonably clearly in extended speech* straddling B1 and B2, thus coming out as B1+. Three descriptors, all concerned with sound recognition, were spread across a range of levels. In the Rasch analysis, the scale produced was very long, from 4.20 logits to -4.60 logits. Because the scale length and slope was very similar to the scale produced in the Swiss research project (North, 2000), straightforward anchoring to the cut-offs between CEFR levels set in that project seemed appropriate. However, only one CEFR item was available as an anchor. Therefore, three more descriptors adapted from Cambridge English Language Assessment criteria for A2, B1 and B2 were also used as anchor items, each anchored to the value of the midpoint of the band of proficiency for the criterion level concerned.

Assessing Performance

The second task was to assess three video performances with 27 descriptors (i.e. excluding the seven descriptors for sound recognition). The videos used were those developed by the Centre international d'études pédagogiques (CIEP) based in France (http://www.ciep.fr/en) in 2008. Respondents chose whether to rate performances in English, French, German or Spanish. There were 872 responses from the 272 participants, but some 200 were excluded for misfit, leaving 667 responses in the data. Two analyses were carried out: one free (unanchored) analysis and one anchored to the North (2000) scale with the same anchor items as in the first task. Again, the scales produced from both analyses was long: from 4.81 logits to -4.96 logits for the unanchored analysis and from 4.96 (high C2) to -4.52 (just below the A1 cut-off) for the anchored one. However, in line with the recommendation of the *Manual for relating assessments to the CEFR levels* (Council of Europe, 2009), a second standard-setting method was also used. This was a simplified form of the Bookmark Method (Council of Europe, 2009, p. 77–81). Each of the five team members independently selected the point for the cut-off between levels on the unanchored scale. The cut-points selected for A2/B1, B1/B2 and B2/C1 by the different team members were identical. Four chose the same cut-point for A1/A2, leaving the C1/C2 cut-point as the only one requiring discussion.

Treatment of Accent

The main issue that emerged in the analysis concerned treatment of accent. The implicit native speaker model of the existing B2 descriptor *Has acquired a clear, natural, pronunciation and intonation* appeared to have instilled the unrealistic expectation that user/learners at the C levels would not have any accent, This B2 CEFR descriptor had actually been calibrated at 2.53, a very high B2+, within the margin of error to the C1 cut-off (North 2000). Yet research has demonstrated that that accent remains a feature of the speech of many people with even a very high level of language proficiency. It is not the "naturalness" of native speech that is essential: it is intelligibility, which is usually not the same thing. Not surprisingly, the statements about accent, intended to be the second part of descriptors at the B and C levels on the *Overall* scale, tended to come out lower than the level intended. One C1 descriptor, however, had been retained intact as a "double barrelled" descriptor with its proviso about accent still attached: *Can articulate virtually all the sounds of the target language; some features of accent retained from other language(s) may be noticeable, but they do not affect intelligibility at all*. This descriptor was calibrated as intended, demonstrating that when statements about accent were added as detail, the effect of rating at a lower level was reduced. Nevertheless, as a result of the responses, the C2 statement about accent was completely rewritten as follows: *Intelligibility and effective conveyance of and enhancement of meaning are not affected in any way by features of accent that may be retained from other language(s).*

Final Decisions

Some descriptors were interpreted in different ways in the two tasks. Deciding which level to finally assign to descriptors was therefore, as always, a question of judgement, based on all the evidence available. There were four pieces of information for most descriptors to guide the decision (two from each task), although the seven descriptors for sound recognition were not used in the assessment task, and therefore had only two information points. In 15 cases, all information points agreed with the intended level and in another five cases, three of the four agreed with the intended level, giving 20 items for which the calibration could be confirmed with confidence. With a further seven items, the information was mixed and for five descriptors it was definitely different to what was intended. One, for *Sound recognition*, was dropped without discussion since it was distributed across all levels. Of the seven others with mixed information, three were accepted as calibrated at a "plus" level, between the criterion levels—but were later dropped from the final scale. Two more were amended slightly in order to make clearer the intended level, and the final two, dealing with accent, were retained where originally intended. The five remaining descriptors, calibrated clearly from what was originally intended, were completely rewritten once the rest of the descriptors on the scales had been completed, in order to provide a coherent scale describing all categories at all levels. The survey had worked well, producing a good, long scale. After some discussion and consultation, however, it was decided to drop the descriptors on sound recognition. The resulting scale is available in the extended version of the CEFR illustrative descriptors (available on the Council of Europe's website). It is presented as an analytic grid in three columns, overall phonological control on the left and the two sub-scales on the right.

CONCLUSION

The need for new descriptors of phonological control aligned with current research in phonological competence, and developed following a methodologically sound process including qualitative and quantitative research, informed the project reported in this article. The aim was to reflect the solid, comprehensive conceptualization of the tenets of phonological competence in the CEFR text. The development, which replaced the original CEFR scale for *Phonological control* and supplemented it with two more scales for *Sound articulation* and *Prosody (intonation, stress and rhythm)* respectively, offered the opportunity for ground-breaking work which hopefully will provide the basis for teachers to include appropriate objectives for phonology in their planning and to develop assessment criteria appropriate to the level(s) concerned.

ABOUT THE AUTHORS

Enrica Piccardo is Associate Professor at OISE – University of Toronto and at the Université Grenoble-Alpes France. A CEFR specialist, she has collaborated with European Institutions on international projects. Her monograph *From Communicative to Action-oriented: a Research Pathways*. (2014) available online is being used in teacher education worldwide. She is the principal investigator of two SSHRC funded research projects, QualiCEFR and LINguistic and Cultural Diversity REinvented (LINCDIRE) and of a Council of Europe funded project QualiMatrix. Her research focuses on assessment, plurilingualism, and creativity in language education. Recent publications include a *TESOL Quarterly* article (2013), a co-edited *The Canadian Modern Language Review* issue (2015), a book chapter on plurilingualism and heritage languages (2014) CUP and one on creativity (2016) De Boek.

enrica.piccardo@utoronto.ca

Brian North is a researcher in language education. He has extensively collaborated with international organizations, particularly the Council of Europe, as a project coordinator and consultant. He spent most of his career at Eurocentres, the Swiss-based foundation that has been an official NGO to the Council of Europe since 1968. In his PhD thesis he developed the levels and descriptors for the CEFR. He then co-authored the CEFR itself, the prototype European Language Portfolio, the Manual for relating examinations to the CEFR, Eaquals' Core Inventories for English and French, and the European Profiling Grid for language teacher competences (EPG). He was Chair of Eaquals from 2005 to 2010. Currently he is coordinating the project to extend the CEFR illustrative descriptors.

bjnorth@eurocentres.com

REFERENCES

Alderson, J. C. A. (1991). Bands and scores. In J. C. A. Alderson and B. North (Eds.) *Language testing in the 1990s*. (pp. 71-86). London: Macmillan.

Baker, A. A. (2011). ESL teachers and pronunciation pedagogy: Exploring the development of teachers‘ cognitions and classroom practices. In. J. Levis & K. LeVelle (Eds.). Proceedings of the 2nd Pronunciation in Second Language Learning and Teaching Conference, Sept. 2010. (pp. 82-94), Ames, IA: Iowa State University.

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

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

Cauvin, E. (2012). Le CECRL et l'évaluation de la compétence phonologique: point de vue à partir d'un corpus sur l'anglais : Evaluer avec le CECRL. *Les Langues modernes 106*(1), 33-40.

Champney, H. (1941). The measurement of parent behavior. *Child Development 12*(2), 131–66.

Council of Europe (2001). *Common European framework of reference for languages: Learning, teaching, assessment (CEFR)*, Strasbourg: Council of Europe available online at: http://www.coe.int/t/dg4/linguistic/cadre1_en.asp

Council of Europe (2009) *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR)*, Strasbourg: Council of Europe http://www.coe.int/t/dg4/linguistic/Manuel1_EN.asp.

Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly, 39*(3), 379-398.

Derwing & Munro (2015). *Pronunciation fundamentals. Evidence-based perspectives for L2 teaching and research*. Amsterdam/Philadelphia: John Benjamins.

Frost, D. & O' Donnell, J. (in press). Evaluating the essentials, the place of prosody in oral production. In J. Volin (ed). *Pronunciation of English as a Foreign or Second Language*

Galaczi, E., Post, B., Li, A. and Graham, C. (2011) Measuring L2 English phonological proficiency: Implications for language assessment. In J. Angouri, M. Daller and J. Treffers-Daller (eds) *The Impact of Applied Linguistics: Proceedings of the 44th Annual Meeting* (pp. 67–72). Downloaded 24.04.17 from: http://www.baal.org.uk/proceedings_11.pdf

Harding, L. (2013). Pronunciation assessment. In C.A. Chapelle (ed.) *The Encyclopedia of Applied Linguistics.* Oxford: Blackwell. . doi:10.1002/ 9781405198431.wbeal0966

Harding, L. (2016). What do raters need in a pronunciation scale? The users' view. In T. Isaacs & P. Trofimovich (Eds.), *Interfaces in second language pronunciation assessment: Interdisciplinary perspectives* (pp12–36). Bristol, UK: Multilingual Matters.

Horner, D. (2010). *Le CECRL et l'évaluation de l'oral: Anglais*. Paris: Belin.

Horner, D. (2013) Towards a new phonological control grid. In E. D. Galaczi & C. J. Weir (Eds.), *Studies in Language Testing: Vol. 36 Exploring Language Frameworks. Proceedings of the ALTE Kraków Conference* (pp. 187-204). Cambridge: UCLES/Cambridge University Press.

Horner. D. (2014) And What about testing pronunciation? A critical look at the CEFR pronunciation grid and a proposal for improvement. in L.M Rupp & R. van den Doel (eds.), *Pronunciation Matters* (pp 109–124). University of Utrecht.

Isaacs, T. (2008). Towards Defining a Valid Assessment Criterion of Pronunciation Proficiency in Non-Native English-Speaking Graduate Students. *The Canadian Modern Language Review/La Revue canadienne des langues vivantes, 64*(4), 555–580.

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

Isaacs, T. and Trofimovich, P. (2012) Deconstructing comprehensibility: identifying the linguistic influences on listener ' L2 comprehensibility ratings. *Studies in Second Language Acquisition 34* (3), 475-505.

Lee, J., Jang, J., & Plonsky, L. (2015). The effectiveness of second language pronunciation instruction: A meta-analysis. *Applied Linguistics36*, 345-366.

Linacre, John M. (2014) *Winsteps: Rasch-model Computer Program*, Chicago: MESA Press.

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

Munro, M. J., Derwing, T. M. (2011). The foundations of accent and intelligibility in pronunciation research. *Language Teaching 44(3)*, 316–327

North, B. (2000). *The Development of a Common Framework Scale of Language Proficiency*, New York: Peter Lang.

North, B and Schneider, G (1998) Scaling descriptors for language proficiency scales. *Language Testing 15*(2), 217–262.

Piccardo, E. (2012). Médiation et apprentissage des langues : Pourquoi est-il temps de réflechir à cette notion ? *ELA : Études de Linguistique Appliquée 167*, 285–97.

Saito, K. (2012). Effects of instruction on L2 pronunciation development: A synthesis of 15 quasi-experimental intervention studies. *TESOL Quarterly, 46*(4), 842-854.

Schneider, G. & North, B. (2000). *Fremdsprachen konnen: was heisst das? Skalen zur Beschreibung, Beurteilung und Selbsteinschatzung der fremdsprachlichen Kommunikationsfahigkeit*, Nationales Forschungsprogramm 33, Chur/Zurich: Wirksamkeit unserer Bildungssysteme, Verlag Ruegger.

Thomson, R. I. & Derwing, T. M. (2014). The effectiveness of L2 pronunciation instruction: a narrative review. *Applied Linguistics 36,* 326–344

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

EXAM-RELATED RESOURCES

Hong Kong Examinations and Assessment Authority (HKEAA). Hong Kong Diploma. of Secondary Education Examination. Available online: http://www.hkeaa.edu.hk/en/hkdse/

HKEAA. Hong Kong Diploma of Secondary Education Examination. English Language Level Descriptors. Speaking Descriptors. Available online: http://www.hkeaa.edu.hk/DocLibrary/HKDSE/Subject_Information/eng_lang/LevelDescriptors-ENG-Speaking.pdf

IELTS SPEAKING: Band Descriptors (public version). Available online: https://www.ielts.org/pdf/SpeakingBanddescriptors.pdf

Educational Testing Service (2014). Independent speaking rubrics/Integrated speaking rubrics. (TOEFL iBT Test). Available online: http://www.ets.org/s/toefl/pdf/toefl_speaking_rubrics.pdf, and scoring guides-Speaking. Available online: http://www.ets.org/s/toefl/pdf/toefl_student_test_prep_planner.pdf

Trinity College London (2015) Integrated Skills in English (ISE) Specifications – Speaking & Listening ISE Foundation to ISE III. Available online: www.trinitycollege.com/resource/?id=6298

Centre international d'études pédagogiques (CIEP) (2008). DVD "Productions orales illustrant les 6 niveaux du Cadre européen commun de référence pour les langues". Available online: http://www.ciep.fr/ressources/ouvrages-cederoms-consacres-a-levaluation-certifications/dvd-productions-orales-illustrant-les-6-niveaux-cecrl

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.

TEACHING ARABIC PRONUNCIATION TO NON-NATIVES: COGNITION AND PRACTICE

Asmaa Shehata, University of Calgary

This study aims to shed light on pronunciation instruction by exploring teachers' beliefs and classroom practices regarding pronunciation teaching. In particular, the study investigates the relationship between Arabic language teachers' beliefs and their current pedagogical pronunciation practices. Data were collected from two groups of subjects (two teachers of Arabic and their 16 students) using different tasks: interviews, class observations, and questionnaires. The result analysis reveals the most frequently used pronunciation techniques in the two different classrooms and how both teachers and students perceived them presenting a brief overview of the way pronunciation is taught in Arabic classrooms illuminating teachers' tendencies and needs.

INTRODUCTION

Research indicates that second language (L2) learners can master L2 morphological and syntactic structures and even gain native-like language abilities (Hinofotis & Baily, 1981). Yet adult language learners are unlikely to achieve nativelike mastery of L2 phonology and some learners experience pronunciation difficulties that cause loss of speech intelligibility (Derwing, Munro & Wiebe, 1998; Zielinski, 2008). After decades of negligence, L2 pronunciation instruction has recently started to receive attention by instructors and researchers alike (Derwing & Munro; 2005; Foote, Holtby, & Derwing, 2011). Several recent studies have explored teachers' cognitions and classroom practices as related to pronunciation (e.g., Baker, 2014; Baker & Burri, 2016; Burri, Baker & Chen, 2017; Sifakis & Sougari, 2005). By and large, most pronunciation studies have mainly focused on teaching English to learners from various language backgrounds, with fewer studies addressing pronunciation issues as related to learners of other languages other than English (Kennedy, Blanchet & Trofimovich, 2014).

Previous research on Arabic language instruction has chiefly focused on grammar and literacy skills (e.g., Ryding, 2013), yet very little is known about Arabic pronunciation instruction. While few studies have explored the perception and production of Arabic segmentals by English speakers (Alwabari, 2013; Shehata, 2015b), a small number have examined classroom contexts and the lack of pronunciation materials that consequently results in students' perception of pronunciation as the most difficult aspect of the Arabic language (Shehata, 2015a; Zouhir, 2013). This paper explores the teaching of Arabic pronunciation and thus it has the added bonus of being about a less-studied context.

Teacher Cognition Research and Pronunciation Instruction

L2 teacher cognition (L2TC) studies have widely investigated teachers' beliefs and knowledge regarding L2 language skills such as literacy (Diab, 2005), grammar (Farrell & Lim, 2005) and assessment (Cohen & Fass, 2001). However, the last decade has witnessed a rapid growth in the L2TC research that addresses pronunciation issues (Baker, 2011; Baker, 2014; Baker & Burri, 2016; Macdonald, 2002). In the ESL context, for example, Macdonald (2002) examined eight ESL Australian teachers' reluctance to teach pronunciation. Findings demonstrated teachers' hesitancy to teach pronunciation owing to their lack of knowledge, training and institutional resources. In a similar vein, Baker (2011) explored five ESL teachers' beliefs and classroom practices regarding the teaching of discourse prosody. Findings reported teachers' lack of confidence to teach certain English prosodic features despite studying pronunciation pedagogy during their graduate education. Similarly, Baker (2014) investigated five ESL teachers' cognitions and pedagogical practices. Findings demonstrated teachers' limited knowledge of communicative pronunciation techniques. More recently, Burri, Baker & Chen (2017) investigated the relationship between participants' teaching experience and learning to teach pronunciation by exploring the cognition development of two groups of teachers (student teachers and in-service teachers) about pronunciation instruction over a postgraduate pronunciation pedagogy course. The results indicated that while student teachers' cognition development was restricted, their awareness about the effectiveness of tactile and kinesthetic techniques developed. Student teachers were challenged to integrate pronunciation into language lessons; however, the beliefs of native English teachers with no pronunciation teaching experience were not developed. In contrast, a small number of studies have focused on teachers' cognition about pronunciation pedagogy in EFL contexts. For instance, Sifakis and Sougari (2005) found that 421 Greek EFL teachers preferred to present the native speaker accent to their students rather than non-native ones. The same tendency was confirmed by other studies (Jenkins, 2007; Timmis, 2002).

To sum up, most L2TC research about pronunciation has focused on English and no much is known about pronunciation instruction in other languages. To date, for example, no research appears to have been conducted to investigate Arabic teachers' cognition (beliefs & knowledge) about pronunciation pedagogy and their practices. Thus, the present study attempts to fill the lacuna in the existing literature by exploring the relationship between Arabic teachers' cognition and classroom practices. To this end, the following two research questions were addressed:

1. What cognitions do Arabic teachers have with respect to techniques of teaching Arabic pronunciation in their classes?
2. Do teachers' beliefs about their own pronunciation teaching practices match observed teaching behavior?

Context and Curriculum

The data was collected from an Arabic language program at a university in the US Midwest. The program has three levels of Arabic: Level 1 (Low & High beginning); Level 2 (low and high intermediate) and Level 3 (low and high advanced) and aimed to help learners to attain proficiency in the four language skills (i.e., reading, writing, listening, and speaking). The current study focused only on two classes: high beginning and low intermediate.

METHODS

Participants

There were two groups of participants. The first group includes two instructors of Arabic who voluntarily agreed to participate in the present study: Magd (High Beginning) and Zahra (Low

Intermediate)[1]. They are native speakers of Arabic and hold Master's degrees in Teaching English to Speakers of Other Languages (TESOL). They have taught different levels of Arabic language courses seven or more years (Teacher1=7 years; Teacher2 = 9 years). The second group included 16 American English speakers who enrolled in their classes. Eight learners were at the novice high proficiency level (i.e., they had spent 8 months learning Arabic) and the other eight students were at the intermediate level (i.e., they had spent 12 months learning Arabic). They were 9 females and 7 males aged from 19-25 years (mean age= 21.4 years). Each student received course credit for participation.

Procedures

The two credit courses focused on developing learners' general language skills and used *Al kitaab fii taalum Al Arabia* textbook. While students in the high beginning course used part I of Al kitaab (Brustad, et al., 2011), their counterparts in the low intermediate course used its second part (Brustad, et al., 2013). Each class met four times a week for 50 minutes for approximately 14 weeks. Summative assessment occurred at the end of each chapter. To explore teachers' cognition and knowledge about pronunciation, several instruments were used to collect data: a questionnaire that teachers completed in week 1 (Appendix A); three semi-structured interviews (SSI) that took place at the beginning (week 1), middle (week 8) and end of the semester (week 14); and four classroom observations (two consecutive classes in week 3 and two consecutive classes in week 11) that were video recorded. The researcher transcribed all the interviews and class observations. Sample of interview questions are presented in Appendix B. Also, students in the two classes completed a questionnaire about their beliefs regarding pronunciation classroom practices that included several sections. However, this paper only reports on the question about the effective teaching techniques that their teachers used and helped them improve their pronunciation.

Data Analysis

Data from the questionnaire, semi-structured interviews and the observed practices were analyzed following the same stages of transcription, segmentation and coding described in Baker (2014) that involved using the qualitative analysis computer program Transana.

RESULTS

Teachers' knowledge of pronunciation techniques

Teachers reported that pronunciation instruction is exhausting and is not well received by students, who prefer practicing vocabulary and grammar. In addition, the two teachers found it too difficult to practice pronunciation due to the lack of materials and their limited knowledge of Arabic pronunciation features other than segmentals (consonants and vowels). While Zahra reported having a pedagogical pronunciation course in her TESOL Master's program that made her confident with English pronunciation instruction, she expressed her unfamiliarity with the Arabic pronunciation features:

> I know nothing about Arabic suprasegmentals and I therefore do not feel confident teaching them. It is my first language that I subconsciously speak without paying

[1] To protect participants' privacy, pseudonyms are used.

> attention to these rules, which I never learned. My knowledge of English pronunciation helped me design different activities for teaching Arabic segmentals though. SSI #2

In contrast, Magd received no education about pronunciation instruction but he confirmed using the textbook pronunciation activities: "Alif Baa[2] is an excellent book that provides various pronunciation activities accompanied with explanations". He also believed that:

> Pronunciation instruction is not necessarily needed beyond the first low level class as by then students would be able to accurately produce the language and thus classroom time should be used to teach other skills such writing and reading, but I only use pronunciation activities in higher level to address students' pronunciation problems when they show up (SSI #1)

They also reported having no experience attending any Arabic-related conferences or pronunciation workshops. In this regard, however, Magd indicated that he attended a six-day workshop about the Arabic language for 21 teachers of Arabic that did not address Arabic pronunciation:

> Last year I attended one of Al Batal[3]'s workshops for training Arabic teachers that presented various techniques for teaching Arabic language skills but pronunciation. I see that pronunciation instruction is not a serious issue that needs much attention. (SSI #1)

When asked about the kinds of activities they used, the two teachers said that they typically used Alif Baa activities (Brustad, et al., 2010) that mainly focus on teaching segmentals. Zahra further explained that no pronunciation materials are presented in the textbooks they used in their current classes:

> Al Kitaab series does not include any pronunciation materials beyond Alif Baa that is used in the beginner low Arabic class. Unfortunately, there are no other sources. Therefore, I occasionally use some of its activities to revive pronunciation in my intermediate class and I sometimes create my own materials to address specific problematic sounds for my students. (SSI #2)

With respect to the first question, data from the classroom observations, teacher interviews and the student questionnaires revealed that Magd and Zahra used a limited number of pronunciation techniques that can be classified, based on Baker's categorization (2014), into three main types: 6 controlled, 2 guided and 3 free. Table 1 shows the techniques used by each teacher as indicated by the interviews, observations and student reports. The two teachers believed that controlled techniques are more beneficial than the guided and free ones in helping students improve their pronunciation. Magd, for instance, said "controlled techniques are meaningful despite their limitations" and he further elaborated,

> Although controlled practice is a stage in a lesson that should be followed by the other two, I find it alone too effective to teach pronunciation. (SSI #3)

---

[2] The textbook used in the first semester of Arabic course

[3] One of the Authors of Al Kitaab series that are widely used in most Arabic courses in North America

Table 1 displays some differences between the data collected through the three research methods. For instance, the interviews showed two more activities for Zahra that were not observed in the four classroom observations (i.e., mutual exchange activity and games) and the observations showed four additional techniques for each teacher that they did not mention in the interviews. Students' questionnaires, however, reported a smaller number of pronunciation activities, especially for Magd where one of his students wrote:

> We are not receiving an actual pronunciation instruction but just bits and pieces every now and then. (SQ)

Although students' reports are compatible with the two teachers' pronunciation teaching approach that aims to teach pronunciation when addressing specific pronunciation problems, the notion of the questionnaire question made students merely focus on those activities they found effective and this explains the small number of activities mentioned by students.

Together, the three data sources confirmed that pronunciation practice was restricted to very few classes in each course where students' opportunities to practice were mainly limited to unassigned textbook exercises that teachers used to address some pronunciation errors. This explains the absence of pronunciation activities in some of the classroom observations (two for Magd and one for Zahra). The data also displayed the two teachers' lack of knowledge of techniques for teaching pronunciation where each teacher used a small number of controlled activities. Furthermore, repetition and production practice activities were found to be the most common controlled drills that the two teachers used.

Table 1

Pronunciation Techniques Used in the Two Classes [4]

Technique	Magd			Zahra		
	TI	O	SQ	TI	O	SQ
Controlled activities						
• Repetition drill activity	X	X	X	X	X	X
• Explanation & examples	X	X	X	X	X	X
• Listening text presentation		X			X	
• Production practice	X	X	X	X	X	X
• Review activity		X			X	
• Testing	X	X				X
Guided Activities						
• Mutual exchange activity	X	X		X		
• Preparation		X			X	
Free Activities						
• Drama		X		X	X	
• Game				X		
• Presentations	X	X	X	X	X	
				X		

Nb. TI= teacher interviews; O= observations; SQ= student questionnaires.

Teachers' Beliefs and Classroom Practices

To answer the second research question, the data from the interviews, observations and the student questionnaires showed the compatibility between teachers' beliefs related to

[4] All categorizations and codes are adapted from Baker's (2014) study

pronunciation techniques and their pedagogical practices except for some activities that teachers did not mention but they were observed in the lessons that were reported earlier. In their discussion of the rationale for using certain techniques, the data showed three different beliefs expressed by the two teachers that are discussed in this section.

Belief 1: Pronunciation chiefly means imitation. For Magd, pronunciation is a mere process of imitation for novel sounds. He believed that "studying pronunciation is less important than studying grammar and vocabulary". This explains his inclination to use controlled activities that he found more useful than others. On the other hand, although Zahra agreed that imitation is an integral part of pronunciation, she believed that "pronunciation includes more than that and it is necessary". Thus, unlike Magd's students who presented memorized dialogues for their final projects, Zahra's students performed novel skits they created.

Belief 2: Pronunciation is exhausting and boring. Magd and Zahra agreed that pronunciation is the most difficult skill to teach because of the lack of time, knowledge, and professional training. Zahra, whose master's degree included a course on pronunciation pedagogy and who relatively used more activities than Magd, believed that the lack of materials is a challenge itself. She said:

> Teaching Arabic pronunciation is very challenging and I sometimes find it too boring to repeat using the same activities. I am not that happy with the materials I use but creating new materials is very exhausting too. (SSI #3)

In contrast, Magd expressed his satisfaction with the Alif Baa textbook, and he found pronunciation instruction "boring and time consuming in comparison with other skills".

Belief 3: attaining a nativelike pronunciation is an illusion. Both Magd and Zahra agreed that comprehensible speech is a feasible goal that Arabic leaners could achieve. For Magd, "nativelike pronunciation does not exist but it was promoted by booksellers and private language centers to deceive language learners and get their money". Zahra also said:

> As long as learners' speech is clear and they can deliver their messages to the listeners, then there is no problem. This is what I usually tell my students especially those who plan to study Arabic abroad. (SSI #3)

Overall, all these quotes were supported with classroom practices and the students' questionnaires**.**

DISCUSSION AND CONCLUSIONS

This study investigated teacher cognitions and practices regarding pronunciation activities in two different Arabic language classrooms. The qualitative analysis of the teachers' interviews and class observations and the student questionnaires demonstrated two important findings. First, the dominance of the controlled activities in the two classrooms at the expense of guided and free activities can be explained in light of teachers' lack of the appropriate pronunciation pedagogy training in the target language. While Magd lacked education in pronunciation pedagogy in general, Zahra's knowledge of English pronunciation pedagogy did not help her teach all Arabic phonological features due to the differences between the two languages. By and large, controlled activities are useful classroom activities whose benefits are displayed in previous research (Couper, 2003); however, they mainly lack the communicative interaction component available in free activities that can help students transfer their classroom learning to real life conversations (Saito & Lyster, 2012). For effective instruction, both controlled and free activities are indispensable. What Arabic instructors need is to learn how and when to use each type of these activities. Second, despite their different beliefs regarding pronunciation instruction and pronunciation activities, the two instructors seemed to emphasize the significance of comprehensible speech in their classrooms. Together, these findings provide preliminarily insights into the nature of Arabic pronunciation instruction that seems to be partially taught in the Arabic classrooms due to the lack of teachers' confidence, appropriate resources and training.

In conclusion, this paper explores the experiences and perceptions of teaching Arabic pronunciation in the US context as seen by two teachers and their 16 students. Despite the study's limitations such as the small number of participants and the absence of teachers' responses to video recordings of their classroom behaviors, I hope it sparks further research on Arabic pronunciation instruction that not only gives a clear picture of the pronunciation pedagogy, teacher cognition and classroom practices but also presents teachers' and learners' needs. What might be of particular interest to explore is teachers' cognitions and actual classroom practices along with students' learning outcomes. Information on this issue is needed to better understand what teaching Arabic pronunciation entails and what best practices are that can inform teacher education programs.

ABOUT THE AUTHOR

Asmaa Shehata is faculty in the School of Languages, Linguistics, Literatures and Cultures at the University of Calgary. Her primary research interest is in second language phonological acquisition and pronunciation instruction.
University of Calgary
Craigie Hall C114. 2500 University Drive NW
Calgary, Alberta, Canada T2N1N4
asmaa.shehata@ucalgary.ca

REFERENCES

Alwabari, S. (2013). *Non-native production of Arabic pharyngeal and pharyngealized consonants*. Unpublished master's thesis, Carleton University, Ottawa.

Baker, A. A. (2011). Discourse prosody and teachers' stated beliefs and practices. *TESOL Journal*, *2*(3), 263–292.

Baker, A. (2014). Exploring teachers' knowledge of second language pronunciation techniques: Teacher cognitions, observed classroom practices, and student perceptions. *TESOL Quarterly*, *48*(1), 136-163.

Baker, A., & Burri, M. (2016). Feedback on second language pronunciation: A case study of EAP teachers' beliefs and practices. *Australian Journal of Teacher Education, 41*(6), 1-19.

Brustad, K., Al-Batal, M., & Tūnisī, A. (2010). *Alif Baa: Introduction to Arabic letters and sounds*. Washington, DC: Georgetown University Press.

Brustad, K., Al-Batal, M., & Tūnisī, A. (2011). *Al-Kitaab fii ta'allum al-'Arabiyya: A textbook for beginning Arabic*. Washington, DC: Georgetown University Press.

Brustad, K. E., Al-Batal, M., & Tūnisī, A. (2013). *al-Kitaab fii ta'allum al-'Arabiyya: A textbook for intermediate Arabic*. Washington, D.C: Georgetown University Press.

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), 108-135.

Couper, G. (2003). The value of an explicit pronunciation syllabus in ESOL teaching. *Prospect*, *18*(3), 53–70.

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

Diab, R. L. (2005). Teachers' and students' beliefs about responding to ESL writing: A case study. *TESL Canada Journal*, *23*(1), 28–43.

Farrell, T. S., & Lim, P. C. (2005). Conceptions of grammar teaching: A case study of teachers' beliefs and classroom practices. *TESL-EJ*, *9*(2), 1-13.

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

Hinofotis, Frances B. & Bailey, Kathleen M. (1981). American undergraduates' reactions to the communication skills of foreign teaching assistants. In J. Fisher, M.A. Clarke, & J. Schacter (Eds.), *On TESOL '80: Building bridges: Research and practice in teaching ESL* (pp. 120-133). Washington, DC: TESOL.

Jenkins, J. (2007). *English as a lingua franca: Attitude and identity*. Oxford, England: Oxford University Press.

Kennedy, S., Blanchet, J., & Trofimovich, P. (2014). Learner pronunciation, awareness and instruction in French as a second language. *Foreign Language Annals*, *47*, 79-96.

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

Ryding, C. K. (2013). *Teaching and learning Arabic as a foreign language: A guide for teachers*. Washington: Georgetown University Press.

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

Sifakis, N. C., & Sougari, A.M. (2005). Pronunciation issues and EIL pedagogy in the periphery: A survey of Greek state school teachers' beliefs. *TESOL Quarterly*, *39*(3), 467-488.

Shehata, A. (2015a). Problematic Arabic consonants for native English speakers: Learners' perspectives. *The International Journal of Educational Investigations, 2*(9), 24-47.

Shehata, A. (2015b). Talker variability and second language word recognition: A new training study. *Journal of Linguistics and Language Teaching*, *6*(2), 209-230.

Timmis, I. (2002). Native-speaker norms and international English: A classroom view. *ELT Journal*, *56*(3), 240-249.

Zielinski, B. W. (2008). The listener: No longer the silent partner in reduced intelligibility. *System*, *36*, 69-84.

Zouhir, A. (2013). Unpacking the teaching and learning practices of Arabic at a major U.S. university. *Journal of Second and Multiple Language Acquisition*, *1*(3), 1-20.

Appendix A: Teachers' Background Questionnaire

1. Age: 2. Highest degree earned: 3. Major:

4. Gender: 5. Native language:

6. What other language(s) do you know?

7. How did you learn them? Explain.

8. How well do you know them?

9. Are you a native Arabic Speaker of Arabic? Yes / No

10. How many years total have you taught Arabic as a foreign language?

11. Were any of your degrees related to the Arabic language?

12. What Arabic-related conferences do you usually attend?

13. How frequently do you attend pronunciation-related workshops or presentations? Do you find them useful?

14. Which Arabic-related journals, magazines or newsletters do you voluntarily read?

Appendix B: Sample Interview Questions

1. What type of Arabic courses you teach?
2. Do you teach pronunciation in your classes? Why? Why not?
3. When you teach pronunciation, what kinds of activities you use? Can you describe them?
4. Which linguistic aspects do you usually teach (segmentals, stress, etc.)?
5. This semester, what features of Arabic pronunciation have you focused on?
6. For each of those features you listed, what activity do you use that is the most helpful for improving students' pronunciation?

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.

LEXICAL ENCODING AND PERCEPTION OF PALATALIZED CONSONANTS IN L2 RUSSIAN

Ala Simonchyk, Indiana University
Isabelle Darcy, Indiana University

Previous studies have investigated the link between lexical encoding and perception by analyzing contrasts that differ in primary features of articulation, e.g., /l/ vs. /ɹ/. The goal of this study was to explore how the lexical encoding of contrasts that differ in the secondary feature of palatalization, e.g., /l/ vs. /lʲ/, was affected by learners' perceptual abilities. The participants in the study were 40 American English learners of Russian and 10 Russian native speakers. Error rates on an auditory word-picture matching task measured learners' ability to encode and retrieve words with the plain/palatalized contrast. Learners' scores on an ABX task assessed their perceptual abilities. Results suggest that learners did not have clearly separated lexical representations for words with palatalized and plain consonants. They accepted most nonwords as possible productions of the target words, especially in the word-final position, whereas Russian native speakers did not. The ability to perceive the contrast between plain and palatalized consonants was found to be helpful in establishing separate lexical representations for words with this contrast among advanced learners, even though it did not guarantee that words with palatalized consonants would be encoded and retrieved as such.

INTRODUCTION

The secondary feature of palatalization, which is phonemic in Russian, is "the superimposition of a raising of the front of the tongue toward a position similar to that for /i/ on a primary gesture" (Ladefoged & Maddieson, 1996, p. 363). There are 15 palatalized consonants in Russian that are paired with plain counterparts. They occur in word-initial, word-medial, and word-final positions, both before vowels and consonants. The most important acoustic cues for palatalization are the first (F1) and second formant (F2) transitions from consonants into subsequent vowels (Halle & Jones, 1959). The F1 of a vowel following a palatalized consonant increases, whereas F2 is high at the beginning of the vowel and then decreases throughout the vowel. The F2 onset of a vowel following a plain consonant is much lower. A vowel that precedes a palatalized consonant has a decreasing F1 and an increasing F2 throughout, but these differences are less salient. The release in palatalized consonants is louder and longer than in plain counterparts (Kochetov, 2002). Depending on the characteristics of a vowel and palatalization status of surrounding consonants, vowel durations can also either increase or decrease (Ordin, 2011).

In perception, learners tend to map plain and palatalized consonants to a single category with different levels of goodness of fit. Plain consonants represent a good fit, whereas palatalized consonants are categorized as a poorer fit to the native category (Rice, 2015). Palatalized consonants are differentiated more easily from plain ones in prevocalic position than in word-final position due to the i-transition or a glide that accompanies palatalization as a cue during the following vowel (e.g., Kochetov, 2002). Under certain conditions, for instance, the perception of palatalized obstruents word-finally, L2 learners do not differ significantly from listeners with no knowledge of Russian. In a study by Lukyanchenko and Gor (2011), L2 learners of Russian with an average of three years of formal instruction in Russian performed similarly to naïve English native speakers on a high-variability AX task that tested the perception of palatalized and plain labials /p – p^j/ and coronals /t – t^j/.

The goal of the current study is to investigate how L2 Russian learners' ability to perceive the contrast between plain and palatalized consonants affects their ability to lexically encode and retrieve words with this contrast and to examine how this relationship unfolds for contrasts that differ in secondary rather than primary features of articulation.

LITERATURE REVIEW

In the first language, the link between lexical encoding and perception is transparent because both domains are reasonably presumed to employ the same phonological categories. In a second language, the interaction between perception and lexical encoding is not as straightforward due to interference of the native categories during processing.

Proponents of the "categories first" approach maintain that the accurate perception of a contrast is necessary for the acquisition of targetlike lexical representations. Inaccurate perception results in single-category assimilation, when two nonnative phonemes are assimilated to the same native category. Consequently, minimal pairs containing that contrast are possibly stored as homophones in the lexicon. Pallier, Colomé, and Sebastian-Gallés (2001) used a repetition-priming paradigm to test fluent Spanish-dominant and Catalan-dominant bilingual speakers of Spanish and Catalan. The participants performed lexical decisions on a list of words containing Catalan-specific phonemic contrasts /e-ɛ/, /o-ɔ/, /s-z/, which do not exist in Spanish. Unlike the Catalan-dominant participants, the Spanish-dominant bilinguals exhibited a repetition effect in the minimal pair condition, whereas their overall performance was similar to that of the Catalan bilinguals. These results indicate that although the Spanish bilinguals mastered the Catalan lexicon, they processed Catalan words with these contrasts as though they were homophones. In a previous study, Pallier, Bosch, and Sebastian-Gallés (1997) found that many early Spanish-dominant bilinguals exhibit a much flatter discrimination function for a continuum between /e-ɛ/ as compared to Catalan-dominant bilinguals, suggesting that they have not established two separate categories despite early and sustained exposure to the contrasts.

The other approach, called "lexicon first," supported by the Direct Mapping from Acoustics to Phonology model (Darcy, Dekydtspotter, Sprouse, Glover, Kaden, McGuire, & Scott, 2012), proposes that the lexical encoding of contrasts is independent of phonetic category formation and can precede it. Learners can use other resources, such as orthography or metalinguistic representations, to establish a lexical contrast. Darcy et al. (2012) examined the acquisition of French vowels /u-y/ and /œ-ɔ/ by American English learners through an ABX and a lexical decision task with repetition priming. Learners' performance on the ABX with /u-y/ was significantly different from that of French native speakers, regardless of their level of proficiency, which means that learners did not yet establish fully robust phonetic categories for the vowel contrasts. On the lexical decision task, intermediate learners exhibited priming effects on the /u-y/ contrast, whereas advanced learners behaved similarly to the French native speakers. These findings suggest that in a lexical task, learners (here, the advanced group) can detect and use more acoustic cues than what they need or use for a nonlexical segmental categorization task. It is possible that at the lexical level a distinction can be made, which may not be implemented in a nonlexical categorization task like ABX. Language experience can help learners overcome spurious homophony and establish separate representations of word forms.

Gor (2014) is the only study to our knowledge that has investigated the perception and lexical encoding of palatalized consonants by heritage learners of Russian and L2 learners of Russian. The participants performed an AXB and a picture-word discrimination task on minimal pairs with /t-t^j/ and /p-p^j/ word-finally, as well as other pairs of consonants in a prevocalic condition /C^{j}V-CjV/. The results of the study showed that L2 learners had accuracy rates of around 70-80% in all three conditions on the ABX task and accuracy rates of 60-76% on the picture-word discrimination task. However, Gor (2014) did not provide a list of minimal pairs that were used in the study, nor was there any mention of whether the words were familiar to learners. The words that form minimal pairs with plain and palatalized consonants in Russian rarely constitute the active vocabulary of Russian learners, especially at lower levels of proficiency. If learners were not familiar with the words in the picture-word discrimination task, then they relied on their phonetic rather than phonolexical knowledge to perform the task. Moreover, the prevocalic condition tested in the study did not represent a clear opposition between plain and palatalized consonants, since prevocalic consonants followed by a palatal /j/ and a vowel tend to be palatalized in Russian, viz. C^{j}jV rather than CjV (Avanesov, 1972).

The current study only used words that were familiar to learners to ensure that participants had already encountered them in spoken and/or written input and established lexical representations for them. Two syllable positions, word-final and intervocalic, were examined because syllable position is expected to have an effect on the lexical encoding of the contrast. The perceptual difference between plain and palatalized consonants in the prevocalic position might be more salient for learners and help them accurately represent words using the palatalization contrast. We used correlational analysis to uncover the relationship between perception and lexical encoding and, consequently, to add to the existing knowledge of the acquisition of contrasts that differ in secondary features of articulation.

RESEARCH QUESTIONS AND HYPOTHESES

The following questions guided the current investigation:

1. Do American English learners of Russian lexically encode words with plain and palatalized consonants separately in L2 Russian?
2. What is the relationship between perception and lexical encoding/retrieval for these learners?

We hypothesize that American learners of Russian should encode plain and palatalized consonants separately, especially at higher levels of proficiency. The perceptual difference between plain and palatalized consonants can alert learners to the existing contrast, especially in intervocalic position. Also, orthographic and metalinguistic knowledge might explicitly direct learners to the differences between plain and palatalized consonants. Palatalized and plain consonants share the same graphemes in Russian, but palatalization is not opaque in spelling. Palatalized consonants are either followed by a letter called the “soft sign” <ь> or a special set of palatalized series vowel letters <и, е, я, ё, ю>.

Regarding the second research question, it is hypothesized that learners' ability to encode palatalized consonants is related to their ability to perceive the distinction. If learners are able to differentiate between plain and palatalized consonants in perception, this reinforces the need to encode the difference. If learners cannot perceive the difference, accurate encoding is still possible if by accurate encoding we mean separate representations for a lexical contrast.

METHOD

Participants

The participants were 40 learners of Russian, all native speakers of American English, from intact classes enrolled in an intensive Russian summer program that offered instruction at nine levels. Enrollment in levels was based on the results of an in-house placement test and previous experience with the language. Participants in each level were tested during their regular Russian Phonetics class. Intermediate participants (9 males, 11 females) aged 19-40 (M = 25.1) included learners enrolled in levels 3-5. Their length of Russian instruction did not exceed 3 years. The advanced group (8 females, 12 males) aged 22-41 (M = 25.9) included learners enrolled in levels 7-9. Their length of Russian instruction was more than 4 years. Ten Russian native speakers (2 males, 8 females) aged 26-42 years (M = 33.3.) served as a control group.

Materials and procedure

Auditory word-picture matching task

An adapted version of the auditory word-picture matching task (AWPM) (Hayes-Harb & Masuda, 2008) was used to examine learners' lexical encoding of words containing plain and palatalized consonants. Stimuli comprised 20 real words contrasting five pairs of coronal consonants, /t-t^j/, /s-s^j/, /n-n^j/, /l-l^j/, /r-r^j/, appearing in word-final or intervocalic position (e.g., [sol̠^{j}] 'salt'). No minimal pairs were used. All words were selected from the textbook *Live from Russia. Volume 2* (Lekic, Davidson & Gor, 1997), which is widely used in first-year Russian courses. In order to ensure that all intermediate and advanced learners would know the words, the materials were piloted with high-beginners who had less than a year of instruction. In addition, participants' familiarity with the target words was evaluated at the end of the testing session. Learners received a list of the target words and fillers in Russian. They were asked to translate the words into English and choose a category that best described their knowledge of each word:

1) I have seen it, I know it, I can use it

2) I saw it, I don't know it

3) I never saw it, I don't know it.

Nine target words (0.9 %) out of 1000 responses were marked as unfamiliar or translated inaccurately. All filler items were familiar to all the participants. The number of syllables, stress and part of speech could not be controlled due to the limits imposed by the vocabulary size of intermediate participants. In word-final position, all target consonants were preceded by the same vowel. In intervocalic position, the vowels that followed the target consonants were the same in words that formed pairs. Twenty test nonwords were created from these real words by replacing the plain or palatalized consonant by its counterpart (e.g., *[sol]). Another 20 control nonwords were created by replacing the target consonant by another consonant differing in primary articulation (e.g., *[som^j]). Twenty additional filler words were selected (e.g., [sumka] ‘purse’), resulting in a total of 80 trials.

During the task, participants saw a picture and were asked to indicate whether the pronunciation of the word they heard was correct and matched the picture by pressing a “Yes” or “No” button as quickly as possible. No written forms were presented. The response timeout was 2000 ms.

ABX categorization task

This task evaluated learners’ abilities to perceptually distinguish the plain vs. palatalized contrast. Ten pairs of CVCVC nonwords were created, in which palatalized consonants alternated with plain counterparts in word-final and intervocalic positions. The vowel /a/ preceded and/or followed the target consonants (e.g., [vatak] – [vat^{j}ak]). Syllables with target consonants were always stressed. Ten pairs of control nonwords with a similar structure contained common contrasts. Each pair was arranged into a triplet (A-B-X) where X was similar to either A or B (e.g., A-[vatak] B-[vat^{j}ak] X-[vat^{j}ak] (X = B). Four counterbalanced orderings were used (ABA, ABB, BAA, BAB), resulting in 40 test triplets. For the control triplets, only two orderings were used, and were counterbalanced for different nonwords; test and control triplets amounted to 60 total trials. Two female Russian native speakers produced the A and B tokens, whereas the X tokens were produced by a male speaker. The interstimulus interval was 500 ms and the response timeout was 2000 ms. Both tasks were administered with DMDX (Forster & Forster, 2003).

RESULTS

Auditory word-picture matching task

Overall, the error rates in all conditions were low for all groups, except in the test nonword condition, where the two learner groups displayed a high error rate (Figure 1). A generalized linear mixed model was run in SPSS 24 on the error rates. The factors group (Russian native speakers, advanced learners, intermediate learners) and condition (target word, test nonword, control nonword, filler) were declared as fixed effects. The factor participant and item were chosen as random effects. Type III tests of fixed effects for error rates revealed that there was a main effect of group, $F(2, 3988) = 30.53$, $p < .001$, condition, $F(3, 3988) = 93.6$, $p < .001$, and an interaction between the two factors, $F(6, 3988) = 14.25$, $p < .001$. Bonferroni post hoc tests showed that intermediate learners with a mean error rate of 82% (95% confidence interval [CI] = 77–85) made significantly ($p = .008$) more errors than advanced learners with a mean error rate of 74% (CI = 69–78) in the nonword condition, when presented with test nonwords */sol/ or */stol^j/ instead of the real words /sol^j/ ‘salt’ or /stol/ ‘table’. The confidence intervals for the two groups’ means on this condition are not overlapping but close, and the mean difference between the two average error rates was 8, (CI = 2–14). Both groups of learners were significantly less accurate than Russian native speakers ($M = 4\%$, CI = 2–8) ($p < .001$ for both groups of learners) on this condition. The mean difference between advanced learners’ and Russian native speakers’ average error rates was 69 (CI = 63–76) and the mean difference between intermediate learners’ and Russian native speakers’ error rates was 77 (CI = 72–83), indicating a robust effect.

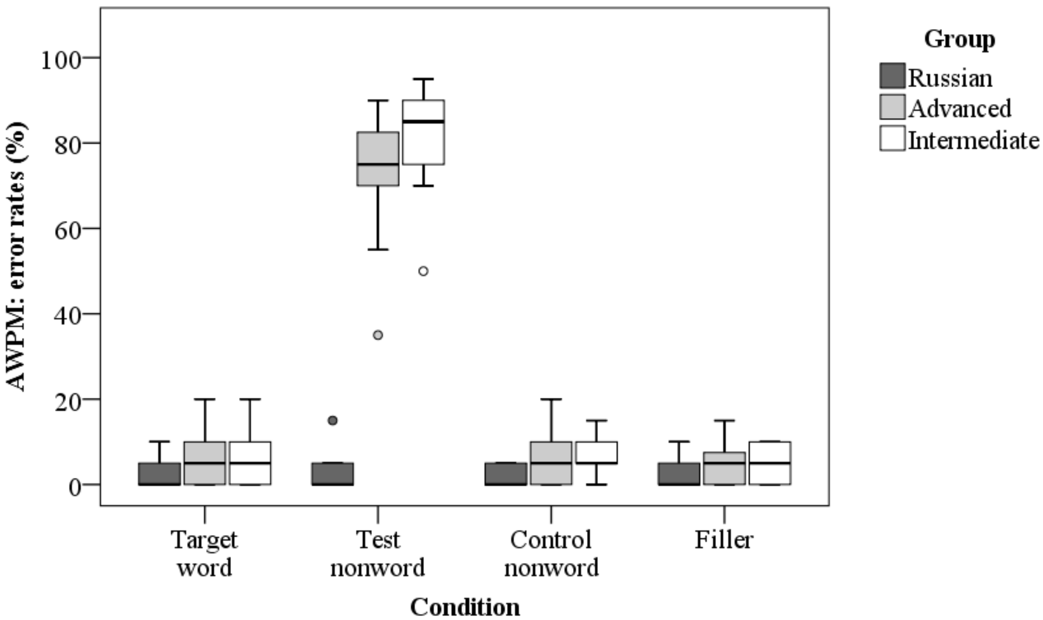


Figure 1. Box plots of error rates for each group and condition. Horizontal lines are medians, boxes show the interquartile range (IQR) representing 50% of the cases, whisker bars extend to 1.5 times the IQR. Outliers (circles) are cases with values between 1.5 and 3 times the IQ range, i.e., beyond the whiskers.

An additional generalized linear mixed model was run on the error rates to examine the effects of syllable position and the palatalization status of the target consonant in the test nonword condition only. The factors group (Russian native speakers, advanced learners, intermediate learners), position (final, intervocalic), and palatalization status (plain, palatalized) were declared as fixed effects. The factor participant was chosen as a random effect. Type III tests of fixed effects for error rates revealed that there was a main effect of group, $F(2, 995) = 56.59$, $p < .001$, palatalization, $F(1, 995) = 4.4$, $p = .036$, and position, $F(1, 995) = 53.68$, $p < .001$, but there were no significant interactions. Additional generalized linear mixed models were run on the error rates to examine the effects of syllable position and palatalization for each group separately. No main effects of syllable position or palatalization were found in the data of Russian native speakers. There was a main effect of position, $F(1, 396) = 20.05$, $p < .001$, in the data of intermediate learners, who made significantly ($p < .001$) more errors in the word-final position (M = 91%, CI = 84–96) than in intervocalic position (M = 73%, CI = 62–82). Confidence intervals do not overlap, and the mean difference was 18, CI = 7–29. There was also a main effect of position, $F(1, 396) = 32.12$, $p < .001$, as well as a marginally significant effect of palatalization, $F(1, 396) = 3.77$, $p = .053$, and a marginally significant interaction between position and palatalization, $F(1, 396) = 3.77$, $p = .053$, in the data of advanced learners. Overall, advanced learners made fewer errors in intervocalic (M = 61%, CI = 52–69) than in word-final position (M = 87%, CI = 81–92). In intervocalic position, the error pattern was modulated by palatalization. Advanced learners made significantly ($p < .001$) more errors by accepting test nonwords with a plain consonant (M = 72%, CI = 61–81), e.g., */z^{j}elonij/ instead of /z^{j}el^{j}onij/ ‘green’, than test nonwords with a palatalized consonant (M = 49%, CI = 38–60), e.g., */xol^{j}odnij/ instead of /xolodnij/ ‘cold’ (see Figure 2). Again, the CIs of the two distributions are not overlapping, and their mean difference was 24, CI = 10–37.

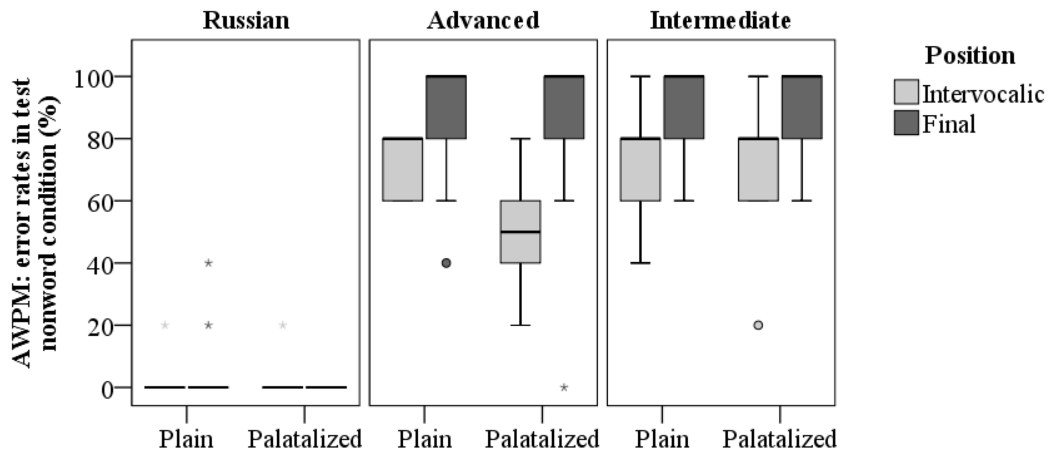


Figure 2. Box plots of error rates on nonwords with plain and palatalized consonants in both positions for each group. See Fig. 1 caption for an explanation of the box plot.

ABX task

As shown in Figure 3, error rates were low in all groups in the control condition, but in the test condition, the learner groups made on average 29% errors. A generalized linear mixed model was run in SPSS 24 on the error rates. The factors group (Russian native speakers, advanced learners, intermediate learners), condition (test, control), and position (intervocalic, final) were declared as fixed effects. The factor participant and item were chosen as random effects. Type III tests of fixed effects for error rates revealed a main effect of condition, $F(1, 2988) = 31.15$, $p < .001$, group, $F(2, 2988) = 8.81$, $p < .001$, and a significant interaction between group, condition and position, $F(7, 2988) = 7.28$ $p < .001$. Bonferroni post hoc tests indicated that learners made significantly ($p < .01$) more errors in the test condition (intermediate: M = 27%, CI = 22–35; advanced: M = 30%, CI = 25–38) than in the control condition (intermediate: M = 4%, CI = 2–7; advanced: M = 7%, CI = 4–13) while Russian native speakers' performance in the test condition (M = 2%, CI = 1–5) was not significantly different from that in the control condition (M = 2%, CI = 1–6). The mean difference between average error rates in the test and control condition for intermediate learners was 23, CI = 19–29, and for advanced learners it was also 23, CI = 19–29. There were no significant differences between the three groups of participants in the control condition. Syllable position had a significant effect ($p < .001$) on learners' performance in the test condition. Both intermediate and advanced learners made significantly more errors in word-final position (intermediate: M = 37%, CI = 29–45; advanced: M = 43%, CI = 35–51) than in intervocalic position (intermediate: M = 18%, CI = 12–23; advanced: M = 17%, CI = 12–22). The confidence intervals between the two positions are not overlapping for either group, indicating a consistent effect. The mean difference between the intervocalic and final position for intermediate learners was 19 (CI = 13–26), and for advanced learners it was 27 (CI = 21–33). There was no statistically significant difference between the two groups of learners.

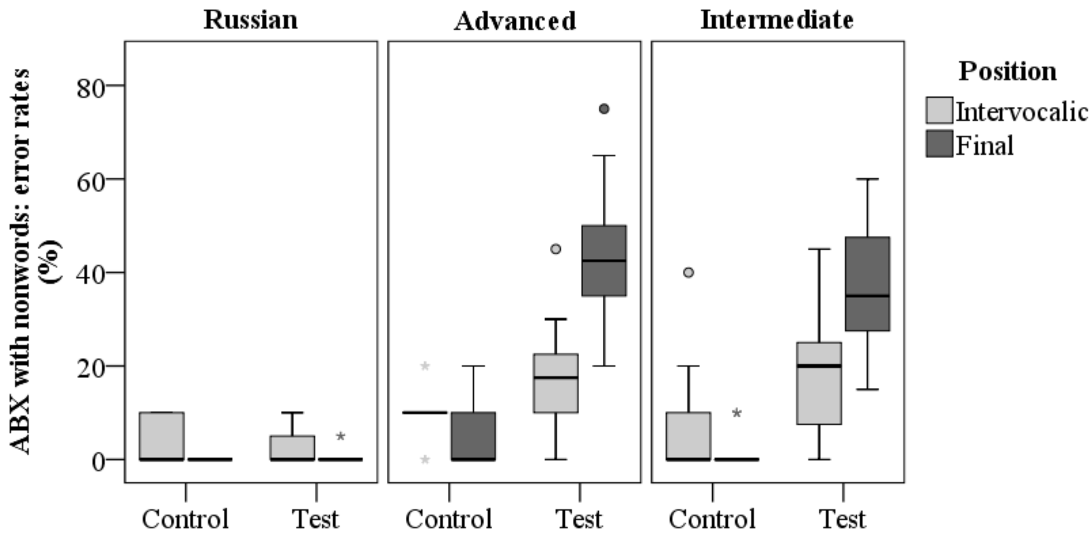


Figure 3. Box plots of error rates for each group, condition, and position. See Fig. 1 caption for an explanation of the box plot.

Correlation

Learners' performance on both tasks was correlated to examine the relationship between perception and lexical encoding. The correlational analysis was performed on the error rates in the ABX task and the error rates in the test nonword condition of the AWPM task for each group separately. For intermediate learners, no relationship was found between their scores on each task, $r(18) = .267$, $p = .256$. However, there was a strong, positive, statistically significant relationship between error rates in both tasks for the advanced group, $r(18) = .657$, $p = .002$. Higher error rates in the ABX were related to higher error rates in the AWPM (Figure 4).

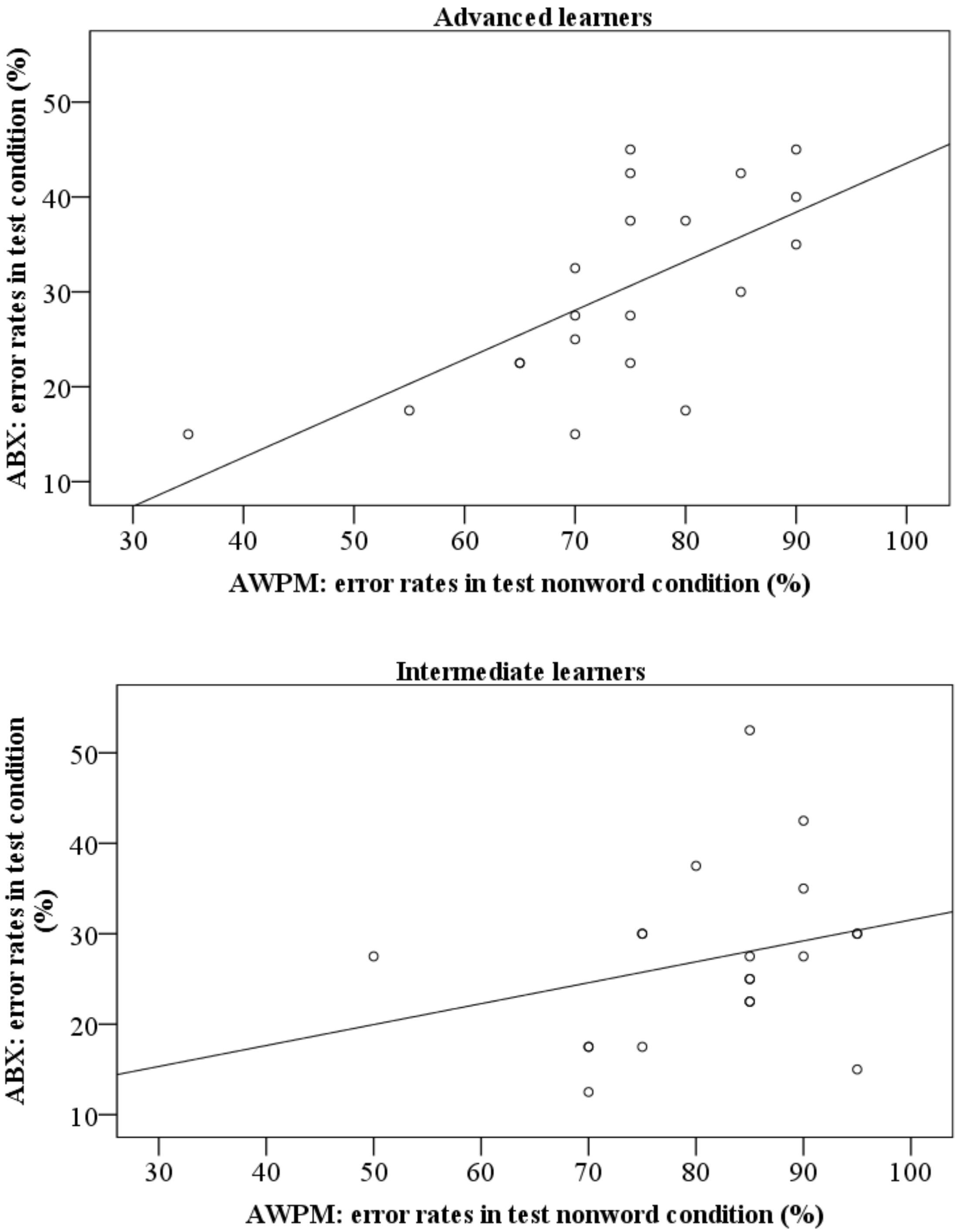


Figure 4. Scatterplots of error rates in both tasks. Upper panel: advanced learners; lower panel: intermediate learners.

DISCUSSION

The goal of the study was to examine the relationship between learners' perceptual abilities and their lexical encoding of words containing plain and palatalized consonants. The first research question asked whether American English learners encoded a clear difference between plain and palatalized consonants in L2 Russian words. The results showed that learners, overall, were not able to encode and retrieve this contrast clearly even in familiar words. All learners mistakenly accepted most test nonwords as correct productions of highly familiar Russian words. Intermediate learners accepted nonwords with either plain or palatalized consonants regardless of syllable position, whereas advanced learners showed an asymmetry in intervocalic position, rejecting test nonwords with a palatalized consonant much more often than nonwords with a plain consonant. Such asymmetry in error rates is reminiscent of findings that rejecting a nondominant (palatalized) category as incorrect in test nonwords is somewhat "easier" than rejecting a dominant (plain) category (Darcy, Daidone & Kojima, 2013). Furthermore, additional acoustic cues to the contrast carried by vowels in intervocalic position, as well as orthographic differences in vocalic graphemes, might have made the difference between target words and test nonwords more salient to learners. Such a strategy, however, might indicate that advanced learners have erroneously encoded the difference in terms of vowels, rather than consonants. In word-final position, since both extra cues are not available, we observed extremely high error rates of 87% for both palatalized and plain test nonwords.

The second research question probed the relationship between perception and lexical encoding. The ABX results showed that learners' perception of the plain/palatalized contrast was not very stable. The two groups showed the same pattern, and they made errors in almost one-third of all trials. The word-final position was more perceptually challenging than the intervocalic position.

The correlational analysis revealed a strong relationship between the perception and lexical encoding of the contrast in the performance of advanced, but not intermediate learners. The learners with the highest error rates in the ABX also had the highest error rates in the AWPM. There was not a single advanced learner with a high error rate in the ABX and a low error rate in the AWPM, which supports the claim that lexical encoding is dependent on learners' perceptual abilities. However, two learners with comparatively low error rates of 15% and 18% in the ABX obtained high error rates (70% and 80%) in the AWPM task. Thus, good perceptual discrimination of the plain/palatalized contrast does not guarantee that words with this contrast are encoded accurately in the mental lexicon. Possibly, despite being able to perceive the difference between plain and palatalized consonants, advanced learners treated Russian palatalized consonants as free variants, and failed to reject most test nonwords in the AWPM task.

In conclusion, our findings suggest a close link between the perception and lexical encoding of contrasts based on the secondary feature of palatalization. The ability to perceive the contrast between plain and palatalized consonants provides a foundation for learners to encode this difference in the mental lexicon. However, perception alone is not enough to guarantee accurate lexical representations of words with a palatalization contrast. The exact reasons for this difficulty are unclear. Future research is needed to uncover the possible influence of orthography and metalinguistic knowledge on the lexical encoding of palatalized consonants in L2 Russian.

ABOUT THE AUTHORS

Ala Simonchyk is a dual PhD student in the Department of Slavic and East European Languages and Cultures and the Department of Second Language Studies at Indiana University. Her research interests include the acquisition of Russian phonology by American English learners and pronunciation instruction.

Indiana University
Department of Slavic and East European Languages and Literatures
Global and International Studies Building
355 N Jordan Ave, 4050
Bloomington, IN 47406
(812) 325-1515
alisimon@indiana.edu

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
Ballantine Hall 866
1020 E Kirkwood Ave.
Bloomington, IN 47405
(812) 855-0033
idarcy@indiana.edu

REFERENCES

Avanesov, R. I. (1972). *Russkoe literaturnoe proiznoshenie.* Moscow: Ucpedgiz.

Darcy, I., Dekydtspotter, L., Sprouse, R., Glover, J., Kaden, C., McGuire, M., & Scott, J. (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.

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

Forster, K., & Forster, J. (2003). DMDX: a windows display program with millisecond accuracy. *Behavior Research Methods, Instruments & Computers, 35,* 116-124.

Gor, K. (2014). Raspberry, not a car: Context predictability and a phonological advantage in early and late learners' processing of speech in noise. *Frontiers in Psychology, 5*, 1449.

Halle, M., & Jones, L. G. (1959). *The sound pattern of Russian*. The Hague: Mouton.

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

Kochetov, A. (2002). *Production, perception, and emergent phonotactic patterns: A case of contrastive palatalization.* New York / London: Routledge.

Ladefoged, P. & Maddieson, I. (1996). *The sounds of the world's languages*. Oxford, UK: Blackwell Publishing.

Lekić, M., Davidson, D., & Gor, K. (1997). *Live from Russia. Volume 2.* (2nd ed.). Dubuque, Iowa: Kendall / Hunt Publishing Company.

Lukyanchenko, A., & Gor, K. (2011). Perceptual correlates of phonological representations in heritage speakers and L2 learners. In N. Danis, K. Mesh, & H. Sung (Eds.), *Proceedings of the 35th Annual Boston University Conference on Language Development* (pp. 414-426). Somerville, MA: Cascadilla Press.

Ordin, M. (2011). Palatalization and intrinsic prosodic vowel features in Russian. *Language and Speech, 54*, 547-568.

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: Concrete exemplar-based vs. abstract lexical entries. *Psychological Science 12*, 445–449.

Rice, H. (2015). *Perceptual acquisition of secondary palatalization in L2: Strengthening the bonds between identification and discrimination through multi-sequence category mapping.* (Unpublished doctoral dissertation). Indiana University, Bloomington, Indiana.

Sturm, J. (2017). Phonetics instruction and the /u/-/y/ distinction in french as a foreign language:a preliminary study. 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. 133-143). Ames, IA: Iowa State University.

PHONETICS INSTRUCTION AND THE /u/-/y/ DISTINCTION IN FRENCH AS A FOREIGN LANGUAGE: A PRELIMINARY STUDY

Jessica Sturm, Purdue University

The distinction between /y/ and /u/ in French, which often proves difficult for native speakers of American English but is important for intelligibility (*rousse* 'redheaded,' n/adj., fem. */russe* 'Russian,'*, poule* 'hen'*/pull* 'pullover sweater') due to its high functional load, has been often studied but not in the context of the effects of classroom phonetics and pronunciation instruction. This study investigates the efficacy of such instruction for learners of L2 French, by comparing progress in the acquisition of the distinction between /u/ and /y/ by university students enrolled in a French phonetics and pronunciation course with a control group of peers not enrolled in phonetics. Learners were recorded at the beginning and end of a semester; tokens of /y/ and /u/ were analyzed using Praat and native speaker ratings and compared to a set of native speaker recordings. Results did not suggest an effect for instruction but support a call for increased early pedagogical intervention in L2 French pronunciation.

INTRODUCTION

A recent surge in research and professional communications on pronunciation in language teaching has found that pronunciation instruction is beneficial (e.g., Lord, 2005, 2008, 2010; Miller, 2012). Building on this renewed emphasis, the current study investigates the benefits of pronunciation instruction for advanced L1 American English learners of L2 French. In particular, this paper is concerned with the distinction between closed rounded vowels /u/ and /y/ by advanced learners (3rd or 4th year university students) of French as a foreign language.

Considerable previous research on L2 pronunciation and phonology acquisition has focused on ESL (e.g., Broselow, Chen, & Wang, 1998; Derwing, Munro, & Wiebe 1997; 1998; Couper, 2006; Rossiter, Derwing, Manimtim, & Thomson, 2010; Saito, 2011; Thomson and Isaacs, 2009;). Hannahs (2007) notes that not much work has been done on L2 phonological acquisition in French. A handful of studies have addressed the production of /y/ and /u/ in L2 French (Darcy, Dekydtspotter, Sprouse, Glover, Kaden, McGuire, & Scott, 2012; Flege & Hillenbrand, 1984; Levy & Law, 2010; Simon, Chambless, & Alves, 2010); these studies focused on learners' production as it stood, rather than on the effects of instruction on learners' ability to produce /y/ vs. /u/ in L2 French, as the current study does.

The /y/ -/ u/ distinction in French

American English (the learners' L1) has only one closed rounded vowel, the back rounded vowel /u/; Flege and Hillenbrand (1984) note that /y/ has no counterpart in English. They also note that the English /u/ is pronounced differently than French /u/; the English /u/ requires the tongue to be farther front than the French one.

The orthography associated with each sound in French ('ou' = /u/; 'u' = /y/) are both associated with /u/ in American English, or in the case of 'ou', also with the mid-high back rounded vowel /o/. As such, the distinction between front rounded /y/ and back rounded /u/ is difficult for native English speakers to recognize, perceive and produce when learning French, which includes both sounds. Flege and MacKay (2004) note that vowels are harder than consonants to perceive in an L2. For the /u/ –/y/ distinction for American English learners of French, Gottfried (1984) points out that both vowels can occur in both open and closed syllables (as opposed to other difficult vowel pairs in French which appear in complementary distribution). Clearly, there is a need to look further at these learners' acquisition of the /y/ – /u/ distinction in French, given the previously reported difficulties and the intelligibility problems that can result.

Functional load and L2 pronunciation

According to King (1967), functional load refers to the contrast between linguistic units, normally phonemes, and can be measured by the number of minimal pairs that can be found for a given pair of phonemes. Munro and Derwing (2006) note that "high functional load errors are predicted to have the greatest impact on listeners' comprehension" (p. 522). Brown (1991) argues for priority to be given to "those conflations [of phonemes] of relatively greater importance" (p. 223). In other words, perceiving and producing phonemes appropriately is essential in oral communication.

In regards to functional load, the /y/ – /u/ distinction is fairly high in French. Meers (2009) lists her target words featuring /y/ or /u/; among 28 words containing /u/ and 26 containing /y/, 25% were minimal pairs. In French, mispronunciation can lead to breakdown in communication, for example: *pull* [pyl] 'pullover sweater'/*poule* [pul] 'hen'; *au-dessus* [od.sy] 'on top of' and *au-dessous* [od.su] 'underneath.' From Meers' (2009) list, we see that the possibility for misunderstanding is strong.

THE CURRENT STUDY

In this study, students enrolled in a pronunciation course were compared with those enrolled in other advanced French courses for their distinction between the high-functional load items /y/ and /u/. Specifically, I examine the following question: Does explicit instruction lead to increased ability to pronounce /y/ and /u/ as two distinct vowels in advanced L2 learners of French? In other words, we are examining only production ability.

METHODOLOGY

Participants

Participants (*N* = 21) were advanced undergraduate learners of French at a large, public Midwestern university. The Phonetics group (*N* = 10) were enrolled in a semester-long phonetics and pronunciation course, taught by the researcher, who is a near-native speaker of French (L1 American English, Ph.D. in French). The class met twice per week, 75 minutes per class session, for 15 weeks. There were four males and six females in the Phonetics group.

Participants in the Phonetics group (enrolled in the class) were not compensated for their time as the tasks involved in the study (a recording at the beginning and the end of the semester) served as part of their course work. There were 14 students in the class, and all consented to have their recordings included in the study, but two were non-native speakers of English; one failed to provide the second recording; and the recordings provided by a fourth student were unanalyzable in Praat (Boersma & Weenink, 2005) due to excessive background noise that made formants impossible to measure. The Control group (*N* = 11) was enrolled in other advanced French courses, but not in the pronunciation class. They were paid $5 each after completing the second recording. There were two males and nine females in the Control group.

Instructional intervention

The pronunciation class was a semester-long course, using an early version of *Sons et sens: La prononciation du français en contexte* (Violin-Wigent, Miller, & Grim 2013). The chapter that included /y/ and /u/ addressed both high and low (closed and open) vowels, beginning with /a/, then /i/, /y/, and /u/. This chapter was covered in one 75-minute class period. Each vowel was introduced in an oral text read twice by the instructor[i], during which students were instructed to answer content questions provided.

During the first reading, students did not look at the text. During the second reading, students were instructed to circle words containing the target sound (in this chapter, each vowel). Follow-up to the second reading consisted of identifying all words with the target sound, answering questions about orthography associated with the sound, and a description (e.g., CLOSED, FRONT, ROUNDED, and ORAL for /y/).

After both /y/ and /u/ were introduced, students completed a discrimination task in which the instructor read a series of words; students marked whether the word contained /y/, /u/, or both. There was an “Expansion” section for each vowel, reviewing orthography and providing pronunciation and transcription practice for each vowel. Homework included transcription and recorded pronunciation exercises, graded and returned by the instructor with comments on pronunciation. These comments used IPA to indicate the students’ actual pronunciation where it deviated from the target pronunciation, along with the target sound. Each chapter (each sound or aspect of pronunciation) proceeded in the same way as the vowel chapter described here.

Students were not held responsible for a given sound or its accurate pronunciation and transcription until that sound had been introduced. The closed rounded vowels were in chapter 14 of 15, so students were only held responsible for accuracy in these vowels for the last two weeks of the semester. By the time that /y/ and /u/ were presented, students were comfortable with the routines of the class and the manner in which feedback was provided.

Data collection

Participant recordings consisted of reading a text provided by one of the authors of *Sons et sens* (Appendix A). Durand and Lyche (2008) note that ‘reading tasks give us systematic access to much of the phonological information we seek…’ (p. 38). A group of native speakers of French from France (*N* = 11) were also recorded, to provide a measure against which both participant groups could be compared. The native speakers were only recorded once, as it was assumed their vowels would not change significantly over the course of a semester. Words containing /u/ and/or /y/ appear in Table 1.

Table 1

Target items with /u/ and /y/.

/**u**/	/**y**/
pourquoi	une
pour	une
tout	du
journée	**sur**tout
sur**tout**	Unis
tous	du
pour	jus
soupe	bu
souvent	confiture
pour	du
nous	une
août	
pour	
ou	
soutient	
nous	

Data Analysis

Recordings were analyzed using Praat (version 5.2.03); measurements were taken for F1, F2, and F3 for each word containing the /u/ or /y/ sound. A Praat script measured each /u/ or /y/ at the midpoint of the segmented vowel. Any F1 measuring higher than 429 Hz[ii], and any F2 measuring higher than 2499 Hz was checked by hand; most of these high measurements turned out to be accurate, but several were inaccurately measured by Praat and the researcher measured these latter cases by hand, at the midpoint of the vowel. If F1 was measured by hand, the corresponding F2 and F3 were also measured by hand; likewise F3 for any re-measured F2. In total, 36% of F1 measurements and 38% of F2 and F3 measurements were checked by hand. Formants were then normalized using the Bark Difference Metric on the University of Oregon NORM Normalization Suite (Thomas & Kendall, 2007).

While all three formants were measured, for the purposes of this paper, the differences between the advancement of both vowels will be analyzed, as the difference between /u/ and /y/ lies in the "frontness," or advancement, of the tongue, as measured by F2, which is transformed via the Bark Difference Method to Z3-Z2.

RESULTS AND DISCUSSION

For all analyses, the alpha level was set at $p = .05$.

Descriptive values for measurements of all three formants, for both vowels at each recording time, appear in Tables 2 and 3 below.

Table 2

Descriptive Statistics, Height, Advancement, and Rounding of /u/ in Barks, Time 1 and Time

Group	N	Avg Height	SD Height	Avg Adv	SD Adv	Avg	SD Round
NS	11	11.29472	0.614125	5.807055	0.809073	5.487667	0.448695
P_1	10	10.65164	0.761196	3.861516	0.85033	6.790128	0.783681
C_1	11	10.40921	0.40719	3.711056	0.511776	6.69815	0.476699
P_2	10	10.80622	0.559946	4.132468	0.902896	6.673753	0.665975
C_2	11	10.26451	0.424311	3.722493	1.567903	6.542019	0.635025

NS = Native Speakers ; P_1 = Phonetics Group Time 1; C_1 = Control Group Time 1; P_2 = Phonetics Group Time 2; C_2 = Control Group Time 2

Table 3

Descriptive Statistics, Height, Advancement, and Rounding of /y/ in Barks, Time 1 and Time 2

Group	N	Avg	SD	Avg	SD Adv	Avg	SD
NS	11	11.6705	0.52321	2.20242	0.22041	9.46812	0.57821
P_1	10	11.1822	0.81983	2.60548	0.52439	8.57676	0.86773
C_1	11	10.9695	0.48112	2.52870	0.30452	8.44087	0.49555
P_2	10	11.3864	0.63233	2.71442	0.89610	8.672	0.77821
C_2	11	10.8412	0.44169	2.51858	3.03289	8.32266	0.76833

To measure the distinction between learners' pronunciation of /u/ and /y/, the difference between their mean advancement for each vowel was calculated and compared. Descriptive statistics for Z3-Z2/y/-Z3-Z2/u/ appear in Tables 4 and 5 below.

Table 4

Descriptive Statistics, Z3-Z2/y/-Z3-Z2/u/, Time 1

Group	N	Mean	SD	SE	Min	Max
NS	11	3.6046	.87562	.26401	2.40	5.04
P_1	10	1.2560	.58639	.18543	.36	2.25
C_1	11	1.1823	.49241	.14847	.12	2.09
Total	32	2.0380	1.32412	.23407	.12	5.04

Table 5

Descriptive Statistics, Z3-Z2/y/-Z3-Z2/u /, Time 2[iii]

Group	N	Mean	SD	SE	Min	Max
NS	11	3.6046	.87562	.26401	2.40	5.04
P_2	10	1.4180	.56080	.17734	.18	2.00
C_2	11	1.2039	.53984	.16277	.62	2.04
Total	32	2.0961	1.29263	.22851	.18	5.04

A one-way Analysis of Variance (ANOVA) showed a significant difference between groups at Time 1, F (2, 29) = 41.5000; $p < .0001$. A Tukey HSD post-hoc analysis revealed the difference to be between each learner group and the NS group (NS/Phonetics, $p < .0001$; NS/Control, $p < .0001$). This suggests that at the beginning of the semester, learners' ability to pronounced /u/ and /y/ as two distinct vowels was equivalent across groups but statistically significantly different from native speakers'. Native speakers, clearly, pronounce /u/ and /y/ as two distinct vowels; their mean distance between Z3-Z2 of the two vowels was significantly greater than both learner groups' mean distance between Z3-Z2 of /u/ and /y/.

An ANOVA of distance between Z3-Z2 of /u/ and /y/ at Time 2 also revealed a significant difference between groups, $F(2, 29) = 45.267$. Again, a Tukey HSD post-hoc analysis revealed the significant difference to be between the NS group and each of the two learner groups and (NS/Phonetics, $p < .0001$; NS/Control, $p < .0001$). This suggests that at the end of the semester, neither learner group pronounced the two vowels in differently. In other words, there was no significant improvement in the Phonetics group after instruction.

DISCUSSION AND CONCLUSIONS

Contrary to expectations, the analyses described above do not suggest an effect for instruction for advanced students on the distinction of pronunciation of the closed rounded vowels /u/ and /y/ in L2 French. Not only was there no significant difference between groups, neither group's pronunciation became more distinct over the course of a semester of French. However, and perhaps more importantly, this study suggests that pronunciation does not simply improve by being in class. The instruction in this particular class was not sufficient for improvement, but based on the body of research on L2 pronunciation, the author believes that instruction in another form could be.

One limitation of this study is the amount of instruction students in the Phonetics group received on the segments in question. As mentioned in the Methodology section, students only studied the /u-y/ distinction during weeks 14-15 of the semester. This observation leads to pedagogical recommendations, which are to begin pronunciation instruction earlier in learners' study of their L2, and to provide repeated instruction on aspects of the L2 that interfere with intelligibility. Anecdotally, pronunciation instruction is largely reactionary and far from systematic in the early years of language learning. Third- or fourth-year phonetics/pronunciation classes such as the one the Phonetics group took are the only consistent instruction that L2 learners (apart from ESL) typically receive. I believe that the results of this study support earlier, systematic intervention, along with the 3rd/4th year phonetics class as a codification of what students have learned as they study the language, with additional exploration of the phonology of the L2. In this same vein, further research may look at a series of lessons on pronunciation of /u/ and /y/ on learners at various stages of competence.

The other limitation of this study is the small *N*. This is a result of the small pool of French majors and minors at the university and the necessarily small classes required to teach foreign languages effectively. Future research could include a larger group of students, or perhaps students at several universities, with more participants and more data points, and therefore more statistical power. However, the caveat there would be ensuring uniformity of instruction over semesters, at different institutions, by different instructors.

This study was designed to look at the effect of a semester-long phonetics class on the distinction between the closed rounded vowels /u/ and /y/ in L2 French. It did so by comparing students enrolled in such a class with peers enrolled in other advanced French classes at the same university. While the results were not as expected, further reflection of the amount of instruction on this particular aspect of French suggest that more systematic, continued instruction on /u/ and /y/ will improve learners' ability to pronounce the two vowels distinctly.

ACKNOWLEDGMENTS

The author wishes to express her appreciation to Anne Violin-Wigent, Jessica Sertling Miller, and Frédérique Grim for the text used in the pronunciation class; Alexander Francis, for guidance above and beyond collegiality in data analysis, revisions and Praat scripts; Pavel Trofimovich, Anita Saalfeld, and Frédérique Grim, for their comments on early drafts of this paper; attendees at the 2012 Association of French Language Studies Conference for their feedback; and Ettien Koffi and the PSLLT Proceedings editors for their comments. All errors that remain are mine.

ABOUT THE AUTHOR

Jessica L. Sturm (Ph.D. 2008, University of Illinois at Urbana-Champaign) is Associate Professor of French and Applied Linguistics in the School of Languages and Cultures at Purdue University, where she also serves as Coordinator of the French Langauge Program. Her research focuses on classroom SLA, specifically the efficacy of instruction on pronunciation in L2 French. She offers graduate courses on teaching L2 pronunciation, acquisition of L2 phonology, and teaching with film, along with undergraduate courses on French phonetics and pronunciation, and French language and culture.

REFERENCES

Boersma, P., & Weenink, D. (2005). Praat: doing phonetics by computer (Version 4.3.01) [Computer program]. Retrieved from http://www.praat.org/

Broselow, E., Chen, S-I, &Wang, C. (1998). The emergence of the unmarked in second language phonology. *Studies in Second Language Acquisition, 20,* 261-280.

Brown, A. (1991). Functional load and the teaching of pronunciation. In Brown, A. (Ed.), *Teaching English pronunciation: A book of readings* (pp. 211-224). New York: Routledge.

Couper, G. (2006). The short and long-term effects of pronunciation instruction. *Prospect, 21,* 1, 46-66.

Darcy, I., Dekydtspotter, R., Sprouse, R. A., Glover, J., Kaden, C., McGuire, M., & Scott, J. (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*(1), 5-40.

Derwing, T. M., Munro, M. J., & Wiebe, G. (1997). Pronunciation instruction for 'fossilized' learners: Can it help? *Applied Language Learning*, *8*, 217-235.

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

Durand, J., &Lyche, C. (2008). French liaison in light of corpus data. *French Language Studies, 18,* 33-66.

Flege, J. E., & Hillenbrand, J. (1984). Limits on phonetic accuracy in foreign language speech production. *Journal of the Acoustic Society of America, 76*, 3, 708-721.

Flege, J. E., & MacKay, I. R. A. (2004). Perceiving vowels in a second language. *Studies in Second Language Association, 26,* 1-34.

Gottfried, T. L. (1984). Effects of consonant context on the perception of French vowels. *Journal of Phonetics, 12,* 91-114.

Hannahs, S. J. (2007). French phonology and L2 acquisition. In D. Ayoun (Ed.), *French applied linguistics* (pp. 50-74). Amsterdam: John Benjamins.

King, R. (1967). Functional load and sound change. *Language, 43,* 831-852.

Levy, E. S., & Law, F. F. (2010). Production of French vowels by American-English learners of French: Language experience, consonantal context, and the perception-production relationship. *Journal of the Acoustical Society of America, 128,* 3, 1290-1305.

Lord, G. (2005). (How) Can we teach foreign language pronunciation? On the effects of a Spanish phonetics course. *Hispania*, *88*, 257-267.

Lord, G. (2008). Podcasting communities and second language pronunciation. *Foreign Language Annals, 41*, 364-379.

Lord, G. (2010). The combined effects of immersion and instruction on second language pronunciation. *Foreign Language Annals*, *43*, 488-503.

Meers, J. (2009). *The acquisition of front rounded and nasalized vowels of French by native speakers of English*. Unpublished undergraduate thesis. University of Calgary, Calgary, Alberta, Canada.

Miller, J. S. (2012). Teaching French pronunciation with phonetics in a college-level beginner French course. *NECTFL Review, 69,* 47-68.

Munro, M. J., & Derwing, T. M. (2006). The functional load principle in ESL instruction: An exploratory study. *System, 34,* 520-531.

Rossiter, M. J., Derwing, T. M., Manimtim, L. G., & Thomson, R. I. (2010). Oral fluency: The neglected component in the communicative language classroom. *The Canadian Modern*

Language Review/Revue canadienne des langues vivantes, *66,* 4, 583-606.

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.

Simon, E., Chambless, D., &Alves, U. K. (2010). Understanding the role of orthography in the acquisition of a non-native vowel contrast. *Language Sciences, 32,* 380-394.

Thomas, E. R. & Kendall, T. 2007. NORM: The vowel normalization and plotting suite. [Online Resource: http://ncslaap.lib.ncsu.edu/tools/norm/]

Thomson, R. I., & Isaacs, T. (2009). Within-category variation in L2 English vowel learning. *Canadian Acoustics/Acoustique canadienne, 37,* 3, 138-139.

Tubach, J. P. (1989). *La parole et son traitement automatique*. CALLIOPE, Collection technique et scientifique des télécommunications. Paris: Masson.

Violin-Wigent, A., Miller, J.S., & Grim, F. (2013). *Sons et sens : La prononciation du français en contexte*. Washington, D.C.: Georgetown University Press.

APPENDIX

Text used for recording[i]

Pourquoi le pain est-il aussi important pour les Français ? Tout simplement parce qu'ils ne peuvent pas envisager une journée, ou même un repas, sans pain. Mais le pain est-il si nécessaire pour manger ? En un mot oui, parce qu'il rehausse les autres aliments.

Pour le petit-déjeûner, rien de meilleur qu'une tartine de pain avec du beurre et de la confiture. A midi, le pain accompagne les salades et surtout soutient le fromage. Entre les deux, il absorbe la sauce de la viande et des haricots. A quatre heures, tous les enfants aiment manger du pain et du chocolat en sortant de l'école. Pour le dîner, de la soupe et du pain suffisent souvent pour un repas équilibré.

Même aux États-Unis, le pain fait partie de la vie. Comment imaginer un sandwich au beurre de cacahuètes sans pain ? Et les hamburgers ?

Phrases

Nous aimons le goût du jus que nous avons bu à Honolulu en août.

La vieille cliente achète des fruits pour Louis. Elle les mange avec lui à minuit dans une ruelle.

Translation

Why is bread so important to the French? Simply because they cannot imagine a day, or even a meal, without bread. But is bread so necessary for eating? In a word, yes, because it enhances other foods.

For breakfast, there is nothing better than a slice of bread with butter and jam. At noon, bread accompanies salads and, above all, supports cheese. Between the two, it absorbs the sauce from the meat and beans. At four o'clock, all children love to eat bread with chocolate when leaving school. For dinner, some soup and bread often suffice for a balanced meal.

Even in the United States, bread is part of life. How could one imagine a peanut-butter sandwich without bread? Or hamburgers?

Sentences

We like the taste of the juice we drank in Honolulu in August.

The old lady client buys fruit for Louis. She eats them with him at midnight in a passageway.

---

[i] The text used in the class, *Sons et sens*, was in development during the semester in which the class was taught and no audio files were available.

[ii] Tubach (1989) gives the following average values for F1 and F2 in hexagonal French: /u/ F1 = 315; F2 = 764; /y/ F1 = 300; F2 = 1750. 429 is 143% of 300; 2499 is 143% of 1750.

[iii] NB: NS were only recorded once; this recording was compared to Time 1 and Time 2 for the learner groups.

[iv] ‘Honolulu’ and ‘hamburger’ were not included in the analysis, as those words are used in English as well and have particularly American associations. Learners may have read them in English or in French; they were not given any specific instructions on those word

Thompson, A. S., & Huensch, A. (2017) Pronunciation attitudes: The role of multilingual status and perceived positive language interaction (PPLI). 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. 144-154). Ames, IA: Iowa State University.

PRONUNCIATION ATTITUDES: THE ROLE OF MULTILINGUAL STATUS AND PERCEIVED POSITIVE LANGUAGE INTERACTION (PPLI)

Amy S. Thompson, University of South Florida
Amanda Huensch, University of South Florida

This study investigated foreign language (FL) learners' attitudes toward improving pronunciation in conjunction with bilingual/multilingual status. The analysis involved both an innovative operationalization of multilingualism, Perceived Positive Language Interaction (PPLI), which is influenced by Herdina and Jessner's (2002) Dynamic Model of Multilingualism, as well as a more traditional definition of multilingualism. The current study explored the relationship between experience with multiple languages and pronunciation attitudes with 195 undergraduate students studying FLs in English-speaking North America. A modified version of the Pronunciation Attitude Inventory (Elliott, 1995a) was combined with data related to bilingual/multilingual status. Quantitative analyses included an exploratory factor analysis (EFA) and subsequent ANOVAs, with the open-ended questions analyzed via a content analysis. The EFA resulted in a three-factor solution: F1 – Lack of NS bias; F2 – Importance of improving pronunciation; F3 – Importance of communication/skills other than pronunciation. The results indicated that F2 was the only factor illustrating group differences, with the multilingual and PPLI participants showing a stronger desire to improve pronunciation in the FLs studied. Ultimately, students with experience with more than one FL have a keener desire to improve their pronunciation, a characteristic that FL instructors can use to their advantage.

INTRODUCTION

Pronunciation is an important language skill, and pronunciation instruction can result in learners becoming more intelligible and comprehensible speakers (Derwing, Munro, & Wiebe, 1997, 1998). To be better prepared to teach pronunciation, instructors need more information about the goals and attitudes of their learners with regard to pronunciation in the classroom. Levis (2015), for example, demonstrated that advanced learners in an English as a Second Language (ESL) context held conflicting/contradictory beliefs about pronunciation improvement. They believed they could 'catch' good pronunciation from native speakers (NS) but did not seek out interactions. Having information about learner characteristics and beliefs can better equip teachers to make informed decisions in the classroom.

While we have some information about learners' attitudes toward pronunciation from an ESL/EFL perspective (Derwing, 2003; Levis, 2015), we know much less about learners of other languages. The majority of work that does exist in this area compares foreign language (FL) learners' attitudes toward pronunciation to their segmental accuracy (e.g., Elliott, 1995a; Shively, 2008) or improvement spanning a certain period of time (e.g., Elliott, 1995b; Kissling, 2014).

For example, to determine factors that contribute to accurate production for intermediate Spanish learners, Elliott (1995a) designed and administered a Pronunciation Attitude Inventory (PAI), which we adapted for use in the current study. Results from Elliott (1995a) indicated that pronunciation attitudes were the most significant predictor of accuracy scores. Elliott (1995b), however, showed that while attitudes moderately correlated with pretest and post-test scores, they did not contribute to a model predicting improvement after one semester of instruction. Similarly, Kissling (2014) found that the best predictor of posttest scores were the pretest scores and also reported significant (albeit small) correlations between two factors: 1. attitudes towards pronunciation and number of university classes completed (r = .24) and 2. attitudes towards pronunciation and extramural language activities (i.e. Spanish use outside of the classroom) measured in hours per week (r = .33). Finally, also investigating learners of Spanish, Shively (2008) found that a U-shaped curve indicated the students with the highest and lowest desire to improve pronunciation had comparable accuracy score profiles (p. 101). In contrast, students who indicated a moderate concern for pronunciation improvement had fairly low accuracy rates in comparison to participants in the other two groups.

Extending pronunciation attitude research beyond the connection between attitudes and improvement/accuracy, Huensch and Thompson (2017) sought to contextualize pronunciation attitudes in U.S. university-level FL classrooms by examining the relationship between attitudes and factors such as language being studied, class level, and extramural language activity. They argued that better understanding attitudes toward pronunciation would allow for knowing how to promote positive attitudes in FL classrooms. The results indicated that the amount of extramural language activity had a strong relationship to positive perceptions about improving pronunciation, especially for learners enrolled in the first two semesters of a FL class, results similar to those in Kissling (2014). However, the authors indicated that further learner variables need to be explored. Considering the fact that a number of students have previous FL learning experience when they enter our classrooms, it is important to understand how attitudes might interact with bilingual/multilingual status. The current work, thus, extends the exploration of the connection between attitudes toward pronunciation and contextual factors by investigating the relationship between pronunciation attitudes and multilingual status of the learners.

What makes a person multilingual? Contrary to popular belief, someone does not need equal levels of competence in all skills to be classified as multilingual (see Gass, Behney, & Plonsky, 2013, pp. 480–481). As such, one of the operationalizations of multilingualism in the current study is those learners who have had experience with two or more foreign languages, regardless of the level of competency achieved. The second operationalization of multilingualism in the current study is the emic perspective of Perceived Positive Language Interaction (PPLI), which classifies a learner as multilingual only if that learner can perceive the positive interactions between the two (or more) foreign languages studied. PPLI was primarily inspired by Kellerman's (1979) work on perceived language distance, with additional theoretical input from the Dynamic Model of Multilingualism (Herdina and Jessner, 2002) and Odlin's (2008) concept of interlingual identification. Essentially, the PPLI framework defines those learners who do not see positive interactions between foreign languages studied as acting similarly to bilingual learners (i.e. learners with one FL) at the cognitive level. In other words, in order to benefit from having studied multiple languages, the learners need to be able to conceptualize the connectivity of the linguistic systems in question. Previous work on PPLI has grouped participants according to their answers on the open-ended question asking about these interactions, and those multilingual learners who see positive interactions are placed in the PPLI group (i.e. seeing the relationship between L2 German and L3 Swedish vocabulary). Those who only had experience with one language other than the L1, those who experienced no interactions between languages, or those who perceived negative interactions (i.e. saying that learning French before Spanish made Spanish grammar harder to learn) are placed in the No Perceived Positive Language Interaction (NPPLI) group (see Figure 1). Further details and examples of the PPLI construct can be found in Thompson (2016).

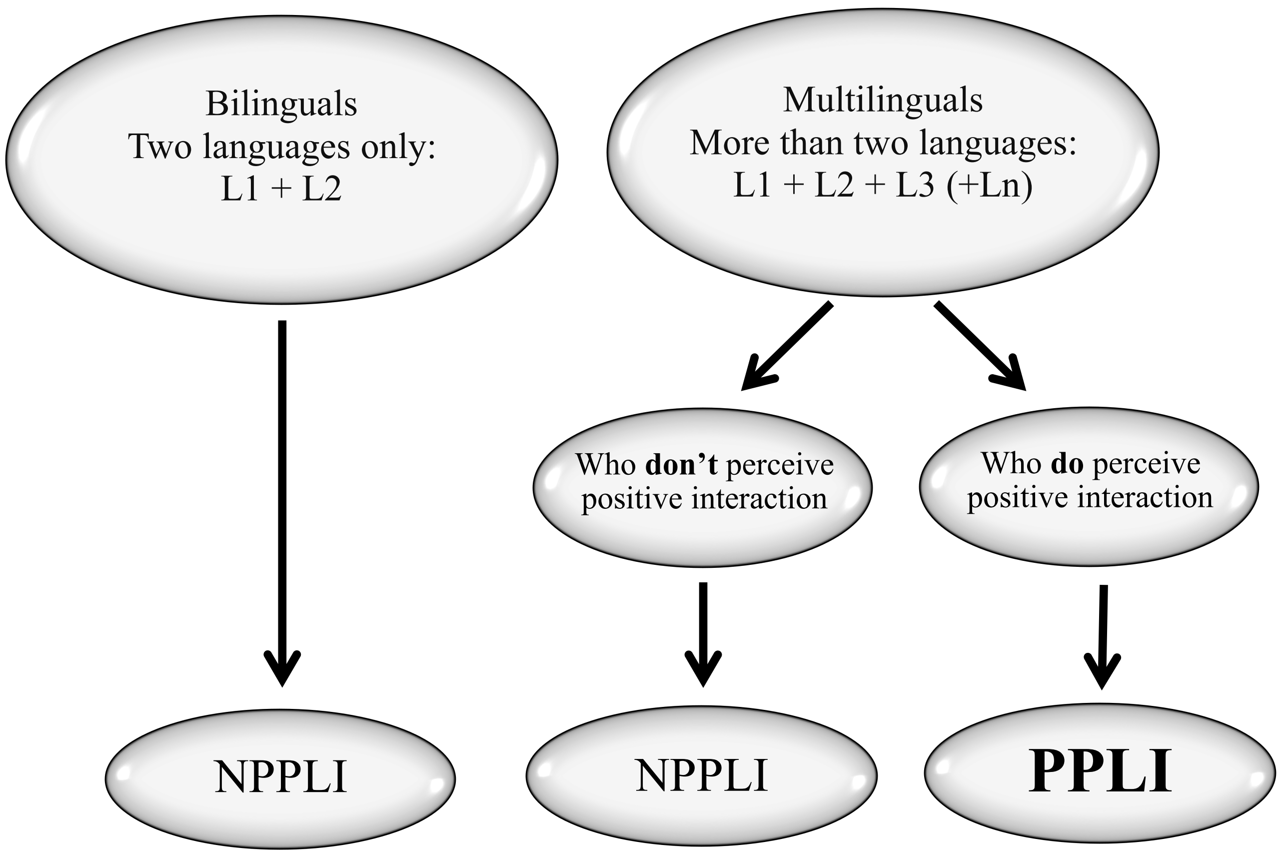


Figure 1. Pictorial representation of the PPLI framework

Several empirical studies have examined the relationship between PPLI and a variety of learner variables, such as language aptitude (Thompson, 2013), motivation (Thompson & Erdil-Moody, 2016), anxiety (Thompson & Khawaja, 2016), and beliefs (Thompson & Aslan, 2015). Additionally, Köylü (2016) examined PPLI and fluency development with Turkish study abroad students, but more work needs to be done regarding the relationship of multilingualism, including the PPLI operationalization, and other learner variables. As such, the impetus of the current study is to examine the relationship of multilingual status and attitudes towards pronunciation.

RESEARCH QUESTIONS

This study is an examination of the relationship between multilingual status and attitudes towards pronunciation in the U.S. university-level FL context. The following are the research questions that guided the study:

1. Do multilinguals have different attitudes towards pronunciation than bilinguals?

2. Do those who perceive positive interactions between foreign languages studied have different attitudes towards pronunciation than those who do not perceive positive interactions?

METHODS

The participants, materials, and procedure used in the current study were described in detail in Huensch and Thompson (2017). Thus a summary of those methods is presented in this manuscript; readers are referred to the aforementioned work for further details, as well as the materials posted on the IRIS website (http://www.iris-database.org/iris/app/home/index – search for the authors' names). Information about the coding of multilingual status is provided in the Data Analysis section.

Participants and Procedures

Participants included 195 university-level FL learners studying a wide variety of foreign languages (see Table 1). The majority of participants were of typical university age (in their 20s) with about 2/3 female and 1/3 male. Participants completed a three-part, online survey that asked about their language learning background, their language learning motivation, and their attitudes toward pronunciation. The survey took about 30 minutes to complete, and those participants who completed the survey were entered into a drawing for one of three $25 Amazon gift cards. As the pronunciation attitudes questionnaire was at the end of the survey, only 180 of the original 195 participants completed this part; thus, 180 participants were included in the analysis for this study.

Data Analysis

As stated in the introduction, two distinct operationalizations of multilingualism were used in the current study. The first operationalization was that those students with experience with two or more foreign languages were classified as multilinguals, no matter the amount of study or the proficiency obtained in the languages. The second operationalization of multilingualism was the PPLI construct, which was defined in the literature review. For the current study, the participants were asked to respond to the following query: "If you have studied other languages in the past, do you think that this has helped or hindered your ability to learn subsequent languages? In other words, do you see interactions (positive or negative) with the languages you have studied? Please provide as many specific examples as you can." The participants were also asked to state the languages with which they saw the interactions and to provide specific examples of these interactions. Twenty percent (randomly chosen) of the open-ended responses were coded for PPLI status by both authors. The Cronbach's Alpha value for inter-rater reliability was .846, indicating strong agreement between the two raters; thus, the remaining 80% were coded by the first author. An example of a statement that would indicate a PPLI grouping is the following: "I have found that by better understanding one romance language/Latin derivative language (Spanish), it was easier for me to begin learning French. Also, as Spanish has limited vowel sounds, it was easier for me to understand the pronunciations of Japanese syllables." The following statement is from a learner who was placed in the NPPLI group: "Because I currently study Chinese intensely I don't have the time to practice French anymore. Additionally the two languages are very different from each other and learning one does not help me the other any better." For the analyses, both operationalizations of multilingualism were used in separate sets of analyses.

Table 1

Participant Demographics

Bi- and Multilinguals	**Bilinguals**	**Multilinguals**
n =180	*n* = 70	*n* =110
NPPLI and PPLI	**NPPLI**	**PPLI**
n =180	*n* =80	*n* =100
Languages **Total** *n* =180	Spanish (68), French (49), German (15), Japanese (12), Chinese (9), English (5), Latin, (5), Arabic (4), Italian (3), ASL (3), Greek (2), Russian (2), Korean (2), Urdu (1)	

RESULTS

Quantitative Analysis

For the analysis, first an Exploratory Factor Analysis (EFA, Maximum Likelihood extraction, direct oblimin rotation) was performed using the 16 pronunciation items (KMO = .852; eigenvalues > 1; item loading values > .3). Details of the EFA procedure can be found in Huensch and Thompson (2017), but the end result was a three-factor solution: F1: lack of NS bias (e.g. not believing that native speakers are inherently more qualified to be pronunciation teachers – items 7 and 16), F2: importance of improving pronunciation (items 1, 2, 3, 4, 5, 6, 8, 11, and 13), and F3: importance of communication/skills other than pronunciation (items 9, 10, 14, and 15). The values from the items that loaded onto each factor were averaged, which resulted in a "factor value" for each participant. The factor values were then used as the dependent variables for group comparisons in the one-way ANOVA analyses in the current study. The effect sizes are reported as η^2 (small: 0.01, medium: 0.059, large: 0.138; Cohen, 1988). All ANOVA analyses were performed twice – once with the multilingual definition of experience with more than one foreign language and again with the PPLI operationalization of multilingualism. The group variables of either “multilingual” or “PPLI” were used as the independent variables in the analyses.

Table 2

ANOVA Results for Multilingual Status

	Bilingual (n = 70)		Multilingual (n = 110)		ANOVA results			
	M	*SD*	*M*	*SD*	*df*	*F*	η^2	*p*
F1: Lack of NS bias	3.53	1.01	3.59	1.17	1, 178	0.116	0.00	.734
F2: Importance of improving pronunciation	4.67	1.01	5.05	0.80	1, 178	8.010	0.04	.005
F3: Importance of communication/skills other than pronunciation	4.23	0.94	4.22	0.90	1, 178	0.001	0.00	.974

Table 3

ANOVA Results for PPLI

	NPPLI (n = 80)		PPLI (n = 100)		ANOVA results			
	M	*SD*	*M*	*SD*	*df*	*F*	η^2	*p*
F1: Lack of NS bias	3.44	1.15	3.67	1.07	1, 178	1.881	0.01	.172
F2: Importance of improving pronunciation	4.60	1.05	5.14	0.68	1, 178	17.488	0.09	.000
F3: Importance of communication/skills other than pronunciation	4.23	1.01	4.21	0.83	1, 178	0.014	0.00	.906

The results of the quantitative analyses indicated that for both operationalizations of multilingualism, there was a significant group difference for F2: Importance of improving pronunciation. With a lower *p*-value and a medium vs. small effect size for the PPLI grouping, the difference of these pronunciation attitudes is amplified when using the emic operationalization of PPLI.

Short Answer Responses

In addition to the quantitative analyses, the responses to the short-answer question about perceived language interactions were analyzed. There was no specific focus on pronunciation in the question posed to the participants; however, some provided answers that focused on pronunciation (27 of 157 responses, 17%). Within the answers that related to pronunciation, there were four themes. Studying other languages in the past:

> (1) generally helped with pronunciation, but with no specific explanation as to how/why (5 comments);
>
> (2) helped because learners had previous experience or larger inventories to draw from (2 comments);
>
> (3) helped because learners could make direct phonological comparisons (10 comments);
>
> (4) hindered (10 comments).

Some of the comments were quite general in terms of previous language experiences helping with pronunciation in subsequent languages. For example, one participant commented, “I think that studying past languages has helped me with learning new languages. I will make connections with some of the words between languages and make connections that way. I also think that it helps with pronunciation.” Other participants found it helpful to have had previous experiences or larger inventories to draw from. For example, one participant stated, “I think studying another language in the past has helped my ability to learn a new language because it has given me experience interacting with a new set of vocabulary rules, grammar, and pronunciation. I see positive interactions between my two language learning experiences.” Another example from this category is the following statement: “It also helps with pronunciations because it increases the amount of morphemes available to me.”

Several participants also made direct phonological comparisons of the foreign languages studied. Some comments were more general, such as, “Learning how to pronounce new sounds in one language aids pronunciation in another.” Some answers compared specific languages, even ones that are seemingly unrelated, such as this statement comparing German and Chinese: “Pronunciation also helps. When learning basic German, I learned how to pronounce ü. English doesn't have this sound, but Chinese has a sound that's very similar.” There were also comparisons between Spanish and Japanese, which are grammatically distinct, yet phonologically similar languages: “Yes, I think learning Spanish before Japanese has helped me with being able to translate from foreign language to English and the other way around. I think Spanish also really helped me understand pronunciation and phonetic sounds.”

Finally, there were also learners who indicated that the one language studied hindered pronunciation in another language. For example, one learner stated that studying French interfered with his Spanish production: “A negative interaction would be how taking French skewed my ability to pronounce Spanish just by looking at it…After taking several more years of French, Spanish is even more difficult to pronounce now, as I have to remind myself that basic, common words like 'de' are different.” Another learner indicated that the different pronunciations of <ll> in Spanish and Italian caused issues: “The similar spelling with dissimilar pronunciation (such as with the word 'pollo') makes learning Italian more difficult.” Although there were some perceived negative interactions in terms of pronunciation in multiple foreign languages, there were more learners who commented on the benefits of the perceived positive interactions with the languages studied.

DISCUSSION AND CONCLUSION

Both operationalizations of multilingualism indicated significant group differences for F2: Importance of improving pronunciation (multilinguals had a stronger desire than bilinguals and the PPLI learners had a stronger desire than the NPPLI learners to improve their pronunciation). The PPLI operationalization of multilingualism demonstrated a more precise division of attitudes toward pronunciation improvement indicated by the smaller p-value and larger effect size, even with the small number of multilinguals who did not indicate a positive perception of interactions between FLs studied (n = 10) being placed in the NPPLI group. This is further evidence of the importance of including the emic perspective of student perceptions when defining multilingual status. Ultimately, these results indicate that experience with multiple languages helps these students enjoy the challenges involved with language learning, including improving pronunciation.

Results from the open-ended question indicated that overall students saw the positive effects on phonological development when having experience with multiple languages with a few participants indicating that learning multiple languages negatively affected their pronunciation in one of the languages. Thus, if students can see positive interactions between languages, the previous language learning experience can help them with subsequent language learning, including seeing parallels with the sound systems that are typically viewed as quite distinct (e.g. Chinese and German). Participants also commented on positive connections between phonologically related languages like Japanese and Spanish potentially without any knowledge of the theoretical aspects of the phonological similarities between the systems.

Why are these results important for language instructors to consider in their classrooms? An increasing number of students enrolled in foreign language classes have experience with multiple languages; thus, instructors should be aware of the language background of their students to more appropriately raise awareness for potential beneficial crosslinguistic interactions. After the instructors have determined the language learning backgrounds of their students (Huensch & Thompson, 2017, has a sample survey to use with students), they can investigate how the students view their past language learning experiences in relation to their current learning experiences. Do they see the learning experiences as separate processes, or do they make connections between them? Thompson (2016) raised the question of the teachability of the perception of positive language interactions. In other words, can instructors help students reap the benefits of their experiences with multiple languages, or is this construct only valid if students come to this realization on their own? After knowing the language background of their students, instructors of less commonly taught languages like Japanese can point out the similarity to other systems (e.g. Spanish). With Spanish being the most commonly studied FL in the U.S. context, encouraging comparisons of FLs can make a less commonly studied language (such as Japanese) more accessible to those students with familiarity with Spanish.

An additional reason for the importance of language instructors considering experience with multiple languages in their classrooms is that students with experiences with multiple languages will have a greater awareness of and a greater desire to work on pronunciation in that language. Even a limited amount of a third language (i.e., one semester) can be beneficial for a variety of reasons, especially in a context where most multilinguals perceive positive interactions between languages, as in the current study.

In conclusion, we need to know more about students' backgrounds to help them with all aspects of language learning, including pronunciation. As the connection between previous language experience and attitudes towards improving pronunciation is a new thread of inquiry, more research needs to be carried out in different contexts to see if the results are replicable. Also, future studies could explore the connection between teacher variables, such as language learning background, and their approach to helping their students with pronunciation, as well as the connection between student and teacher beliefs surrounding the need for explicit attention on pronunciation during class time. The connection between learner variables and attitudes towards improving FL pronunciation is an area ripe for future research.

ACKNOWLEDGMENTS

We would like to thank the participants for participating in this study.

ABOUT THE AUTHORS

Amy S. Thompson (Ph.D. Michigan State University, 2009) is an Associate Professor of Applied Linguistics and currently the Associate Department Chair in the Department of World Languages at the University of South Florida. She is also currently the graduate director for the Ph.D. program in Linguistics and Applied Language Studies (LALS), teaching a range of graduate level theoretical and methodological courses in applied linguistics. Her primary research interests involve Individual Differences in Second Language Acquisition and the interaction of these IDs and multilingualism. Examples of her research can be found in journals such as the *Modern Language Journal, TESOL Quarterly, Foreign Language Annals*, and the *International Journal of Bilingual Education and Bilingualism.*

Amanda Huensch (Ph.D. University of Illinois at Urbana-Champaign, 2013) is an Assistant Professor of Applied Linguistics in the Department of World Languages at the University of South Florida, where she teaches courses in second language acquisition, pronunciation pedagogy, and applied linguistics. Her research examines second language speech development in and outside of the classroom, including the relationship between perception and production, pronunciation attitudes of classroom foreign language learners, and fluency development during study abroad. Her most recent work has been published in the *Journal of Phonetics*, the *Journal of Second Language Pronunciation*, and *Applied Psycholinguistics*.

REFERENCES

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Derwing, T. M. (2003). What do ESL students say about their accents? *Canadian Modern Language Review*, *59*, 547-567. doi: 10.3138/cmlr.59.4.547

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

Derwing, T. M., Munro, M. J., & Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning*, *48*, 393-410. doi: 10.1111/0023-8333.00047

Elliott, A. R. (1995a). Field independence/dependence, hemispheric specialization, and attitude in relation to pronunciation accuracy in Spanish as a foreign language. *The Modern Language Journal*, *79*, 356-371. doi: 10.1111/j.1540-4781.1995.tb01112.x

Elliott, A. R. (1995b). Foreign language phonology: Field independence, attitude, and the success of formal instruction in Spanish pronunciation. *The Modern Language Journal*, *79*, 530-542. doi: 10.1111/j.1540-4781.1995.tb05456.x

Gass, S., Behney, J., & Plonsky, L. (2013). *Second language acquisition: An introductory course* (4th ed.). New York: Routledge.

Herdina, P., & Jessner, U. (2002). *A dynamic model of multilingualism: Perspectives of change in psycholinguistics* (Vol. 121). Clevedon: Multilingual Matters.

Huensch, A., & Thompson, A. S. (2017). Contextualizing attitudes toward pronunciation: Foreign language learners in the United States. *Foreign Language Annals,* doi:10.1111/flan.12259

Kellerman, E. (1979). Transfer and non–transfer: Where are we now? *Studies in Second Language Acquisition, 2*, 37-57. doi: 10.1017/S0272263100000942

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 Languages Review, 70*, 532-558. doi: 10.3138/cmlr.2161

Köylü, Z. (2016). *The influence of context on L2 development: The case of Turkish undergraduates at home and abroad* (unpublished doctoral dissertation). University of South Florida, Tampa, Florida.

Levis, J. (2015). Learners' views of social issues in pronunciation learning. *Journal of Academic Language and Learning*, *9*, A42-A55.

Odlin, T. (2008). Cross-linguistic influence. In C. Doughty & M. Long (Eds.), *The handbook of second language acquisition* (pp. 436-86). Malden, MA: Blackwell.

Shively, R. L. (2008). L2 acquisition of [β],[ð], and [ɣ] in Spanish: impact of experience, linguistic environment, and learner variables. *Southwest Journal of Linguistics*, *27*, 79-115.

Thompson, A. S. (2013). The interface of language aptitude and multilingualism: Reconsidering the bilingual/multilingual dichotomy. *Modern Language Journal, 97*, 685–70. doi: 10.1111/j.1540-4781.2013.12034.x

Thompson, A. S. (2016). How do multilinguals conceptualize interactions among languages studied? Operationalizing perceived positive language interaction (PPLI). In L. Ortega, A. Tyler, & M. Uno (Eds.) *The Usage-based study of language learning and multilingualism* (pp. 91–111). Georgetown: Georgetown University Press.

Thompson, A. S. & Erdil-Moody, Z. (2016). Operationalizing multilingualism: Language learning motivation in Turkey. *International Journal of Bilingual Education and Bilingualism, 19*, 314-331. doi: 10.1080/13670050.2014.985631

Thompson, A. S. & Khawaja, A. (2016). Foreign language anxiety in Turkey: The role of multilingualism. *Journal of Multilingual and Multicultural Development, 37*, 115–130. doi: 10.1080/01434632.2015.1031136

Thompson, A. S. & Aslan, E. (2015). Multilingualism, perceived positive language interaction (PPLI), and learner beliefs: What do Turkish students believe? *International Journal of Multilingualism, 12,* 259–275. doi: 10.1080/14790718.2014.973413

White, D., Gananathan, R., Mok, P. (2017) Teaching dark /l/ with ultrasound technology. 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. 155-175). Ames, IA: Iowa State University.

TEACHING DARK /l/ WITH ULTRASOUND TECHNOLOGY

Donald White, The Chinese University of Hong Kong
Richard Gananathan, The Chinese University of Hong Kong
Peggy Mok, The Chinese University of Hong Kong

This study examined the efficacy of ultrasound technology for teaching L2 English "dark /l/", [ɫ], to L1 Cantonese students. In a pre-test reading, eight Hong Kong students, aged 15-16 (mean age: 15 years, 9 months), were scanned with ultrasound while reading short token sentences containing word-final [ɫ] tokens in three vowel contexts: /iː/, /ɔ/, and /uː/. Students were then divided into two groups of four, and given a short lesson on the articulation of [ɫ]. The two lessons were identical in every respect except for one: Group 1 received visual feedback from the ultrasound scanner, but Group 2 did not. Finally, in a post-test immediately after the lesson, the students were scanned reading the token sentences a second time. Results from the pre-test indicated that all students except for one in group 2 articulated [ɫ] with a back tongue gesture, but no subsequent front gesture. The resulting sound was vocalized, and more akin to a back vowel than [ɫ]. In the post-test, 3 out of 4 students from Group 1 added a front tongue gesture to [ɫ]; however, in Group 2 there were no major differences between the pre-test and post-test results. The changes in Group 1 occurred across vowel contexts, with somewhat different effects from vowel to vowel.

INTRODUCTION

This study is a preliminary investigation into the efficacy of ultrasound technology in second language (L2) speech instruction. Eight native (L1) Cantonese speakers participated in a pre-test/post-test experiment that was carried out in a Hong Kong secondary school. The target English sound was syllable-final English dark /l/ (henceforth [ɫ]), as in the word "hall". The difficulty of [ɫ] for Cantonese L1 speakers derives from its complex articulation, which involves a back tongue gesture followed rapidly by a front tongue gesture. Additionally, [ɫ] is not a part of the Cantonese phonological inventory (Chan & Li, 2000). For these reasons, Cantonese L1 speakers tend to vocalize [ɫ], and articulate it as a back vowel such as /uː/. In order to test the effectiveness of ultrasound as a method for teaching pronunciation, the eight participants took part in a three-step process: first, they were scanned reading [ɫ] tokens in carrier sentences; then they were given a lesson on the proper articulation of [ɫ]; and, finally, they were scanned reading the carrier sentences a second time.

Previous L2 Instruction using Ultrasound

The use of ultrasound for pronunciation instruction is a relatively new form of articulatory feedback for learners of L2 speech. Technologies employed to provide articulatory feedback have included verbal instructions (Catford & Pisoni, 1970), mirrors (Firth, 1987), and more complex technologies. In the past, however, higher complexity of feedback has often led to intractable difficulties. For example, Truby's (1959) X-rays, which provide a stunningly clear view of the articulators in use, are inappropriate for L2 learning, as prolonged use would endanger the lives of the learners. Overcoming many of these problems, ultrasound scans have emerged as a speech-learning tool that is both technologically complex and feasible for practical use. This technology is harmless, relatively non-invasive, and while expensive, it is becoming much more affordable with the passage of time (Gick, Bernhardt, Bacsfalvi, & Wilson, 2008).

In their most common phonetic application, ultrasound scans provide a mid sagittal view of the tongue that excludes the root and tip. For this reason, some speech sounds are more conducive to successful ultrasound scans than others. In previous ultrasound studies of L2 speech instruction, therefore, the most common phones investigated have been liquids and glides, whose correct articulation corresponds to specific tongue contours in the blade and body of the tongue.

Several previous studies have used ultrasound to teach the articulation of [l] and [ɹ] to Japanese learners of English. First, Gick, Bernhardt, Bacsfalvi, & Wilson (2008) conducted a pilot experiment with three recently arrived Japanese immigrants to Canada. Ultrasound scans were performed on these participants while reading [l] and [ɹ] tokens in word-initial, word-medial, and word-final contexts. After these scans, the pronunciation difficulties of each participant were identified, and the ultrasound images and scanner were used in a variety of ways to demonstrate correct articulation. A post-lesson scan then showed that all three participants were able to produce [l] and [ɹ] more accurately in the three contexts

Using a similar design, Tsui (2012) found that six Japanese adults improved their pronunciation with the help of ultrasound scans. In this case, the lessons were more detailed, and there was up a two-week period between the pre-test and post-test. Tsui (2012) found that the improvement was generally more robust for articulation of [l] although there was improvement for [ɹ] as well.

In another study of 10 newly arrived Japanese immigrants to Canada, Tateishi (2013) had similar results to Tsui (2013): the production of [l] improved more than that of [ɹ] after a training session using ultrasound. However, this study also tested the ability of the subjects to distinguish [l] and [ɹ], and no significant improvement was evident in perception.

Although L2 liquids and glides have been investigated a number of times in ultrasound studies, very little work has examined the English syllable-final variant, [ɫ], which is the focus of the present study. One exception is King & Ferragne (2015), a work in progress that is using ultrasound to examine [ɫ] production among L1 French-English L2 learners. Thus far, ultrasound has been employed solely as a diagnostic tool to establish an articulatory difference between native English speakers and French L1-English L2 speakers in their pronunciation of [ɫ]. Currently, they are designing the next phase of their study, which will include ultrasound training sessions.

The present study adopts an approach similar to those above, but two elements distinguish it from previous work. First, no previous ultrasound speech training studies have focused exclusively on the acquisition of [ɫ] in Cantonese L1 subjects. Second, to the best of our knowledge, ultrasound training has never been used before as an instructional method in Hong Kong.

METHOD

Participants and Lesson

Among the eight participants in the current study, there were four males and four females. These subjects were recruited from the same Form 4 (10^{th} grade) class of a secondary school in Lam Tin, Kowloon. After Form 3, the students at this school are streamed according to their English ability into one of five Form 4 classes. All of the students in the present study were from the second-best class, meaning that their level of spoken English was roughly the same.

The eight students were then divided into two groups of four. This division was based on two criteria: equity of male to female ratios, and parity of English speaking ability. Each group, therefore, comprised two boys and two girls, and the average L2 speaking proficiency for each group was roughly the same. To determine these levels, the first term speaking examinations results of the participants were compared. The speaking exam is an eight-minute group discussion among four students, which is graded by a single examiner. There are four categories in the marking scheme for this exam: 1. Pronunciation and Delivery; 2. Communication Strategies; 3. Vocabulary and Language Patterns; and 4. Ideas and Organization. Each category is worth 25 percent of the total mark. Table 1 shows the results for each group member as well as the mean result and standard deviation for each group.

Table 1

Term 1 Speaking Exam Results of Participants (Total Marks: 28)

Group 1	**Group 2**
Boy 1 – 11	Boy 2 – 12
Boy 4 – 14	Boy 3 – 13
Girl 3 – 14	Girl 1 – 14
Girl 4 – 13	Girl 2 – 14
Mean (SD) – 13 (1.414)	Mean (SD) – 13.25 (0.957)

For the pre-test and post-test, students were scanned with ultrasound and recorded while reading sentences with embedded [ɫ] tokens in three vowel contexts: [iː], [ɔ], and [uː]. The three sentences were as follows:

1. The meal heater was broken.
2. The hall model was very nice.
3. The pool food cannot be eaten.

These semantically unusual sentences were composed by the authors according to a single criterion: minimal tongue movement in the vicinity of the [ɫ] token. In order to minimize tongue movement, the onset of each token's syllable could be either a labial or a pharyngeal. This restriction also applied to the onset of the syllable following the [ɫ] token. Additionally, the nucleus of the following syllable was the same as the nucleus preceding the [ɫ] token. All of these restrictions ensured that there would be minimal co-articulatory influence on the tongue shape of the relevant tokens. Each sentence was read a total of 10 times by each participant: five times in the pre-test and five in the post-test.

Between the two tests, each group received a 30-minute lesson on how to properly articulate [ɫ]. The lesson for each group was identical except for one key feature: Group 1 received visual feedback from the ultrasound scanner, but Group 2 received no such feedback. The time between the pre-test and the post-test was approximately 1.5 hours. (Please see the Power point slides from the lessons in Appendix 1, which represents the lesson for Group 1; Group 2's lesson contained the same slides except for the two that refer to ultrasound.)

Ultrasound Scanning

During the pre-test and post-test, a stabilization helmet was employed to stabilize the heads of the participants.

Figure 1: Stabilization helmet (left) and ultrasound apparatus (right)

Helmet stabilization was not employed in the ultrasound lesson in order to facilitate scanning each student in a relatively short amount of time. Stabilization was not required for this part because the data collected during the lesson were not analyzed.

The pre-test and post-test ultrasound data were analyzed using the Articulate Assistant Advanced (AAA) software package (Articulate Instruments Ltd., 2012). The [ɫ] segments were annotated according to their acoustic boundaries, and a single ultrasound frame per token within this boundary was chosen as the representative ultrasound frame for that token. Ideally, articulatory criteria were used to select this representative frame, wherein the frame with the highest rightmost visible edge of the tongue was chosen. In most cases, this would mean the blade, and likely the tip as well, would be raised. In some cases, however, it was difficult to use these criteria to select the representative frame, because the position of the tongue did not change greatly throughout the acoustic boundary of the segment. In these cases, a frame close to the end of the segment was selected.

An SS-ANOVA (Gu, 2002) was conducted on the results to compare the tongue spline shapes of the data from the pre-test and post-test. Ideally, an SS-ANOVA would be conducted to compare the pre-test and post-test tongue splines directly;

however, as the pre/post-test data were taken from separate ultrasound recording sessions, a direct comparison is not permitted because of possible variation in the position and angle of the ultrasound probe. Rather than a direct comparison between the pre-test and post-test splines, therefore, two SS-ANOVA comparisons were conducted: one within the pre-test data, and one within the post-test data. In each data set, the highest point of the front-most, visible part of the tongue during the acoustic duration of the [ɫ] was compared with the lowest point during the first half of the [ɫ] or immediately prior to it. (The demarcations between nuclei and [ɫ] were determined from spectrograms of the data.) In this way, the difference between the highest point and the lowest point in the pre-test can be compared with the difference between the highest point and the lowest point in the post-test. If the highest and lowest points were overlapping or very similar, it would suggest that the [ɫ] is likely to be vocalized, and there would not likely be tongue-tip raising. If the highest and lowest points were separated from each other, it would suggest tongue-tip raising during the highest point, indicating a correct production of [ɫ]. In other words, a stronger distinction between the highest and lowest points in the post-test would indicate improvement in the articulation of [ɫ]. When produced correctly, [ɫ] generally starts off with a back tongue gesture and only near the end is there a tongue tip raising gesture. Cases in which the first half of the [ɫ] has a higher point than the second half did sometimes arise when the preceding vowel is an [iː], and seemed to be due to the influence of coarticulation from the vowel, which is followed by a lowering of the tongue tip for a vocalized coda [ɫ]. Because of this phenomenon, the “lowest point” could come from only the first half of the [ɫ] or before, and the highest point from some point following it anywhere in the [ɫ]. In this way, only if the [ɫ] is produced correctly will the SS-ANOVAs show a sharp difference between the highest point and lowest point, but if coarticulation from a high vowel causes the highest point to precede the lowest point within the duration of the [ɫ], these cases will not be considered to be exemplary productions. To complement these data, the two native English-speaking authors have also judged the acceptability of each of the tokens.

RESULTS

Figures 2-7 show the SS-ANOVAs comparing the pre-test and post-test tongue splines of each individual participant (labelled a-d: “a” include the pre- and post-test splines for participant 1, b for participant 2, and so on). The pre-test SS-ANOVAs are on the left and the post-test SS-ANOVAs are on the right. In the SS-ANOVAs, the pink splines represent the lowest point of the front-most visible part of the tongue during the first half of the acoustic duration of the [ɫ] or before (point A), and the cyan splines represent the highest point of the front-most visible part of the tongue during the acoustic duration of the [ɫ] (point B). Individual tracked ultrasound points are reported using dots. The line in the center of each shaded area represents the mean tongue position for the five tokens. The shaded area above and below these lines is the standard error for each mean. Where the pink and cyan bars overlap, there is no statistically significant difference between the pre-test and post-test tongue spline contour. Where the pink and cyan bars do not overlap, there is a statistically significant difference between the pre-test and post-test results.

The results for the [ɔ] vowel are the most clear. In all cases, the pre-test results (on the left) show either a completely overlapping point A & point B, or the difference between point A & point B is small. The post-test results (on the right) show a higher separation between point A & B. The results for the [iː] condition are also very clear.

In Figure 3 a-c, it is clear that the pre-test results (left) all have overlapping point A and point B SS-ANOVAs, but in the post-test results (right) they are no longer overlapping near the tip of the tongue. In the [uː] results, the differences between the pre- and post-test splines are not as great; however, in Figure 4c there is a clear improvement in the post-test. Perhaps because of [uː]'s similarity to the vocalized coda [ɫ], it was harder for the students to learn to raise the tongue in this condition. Another possibility is that because [uː] is a high vowel, even if the tongue tip was raised, point A may still be high due to coarticulation from the [uː]. It is not clear why this effect did not seem to happen for [iː], however.

Pre-test Post-test

a. b. c. d.

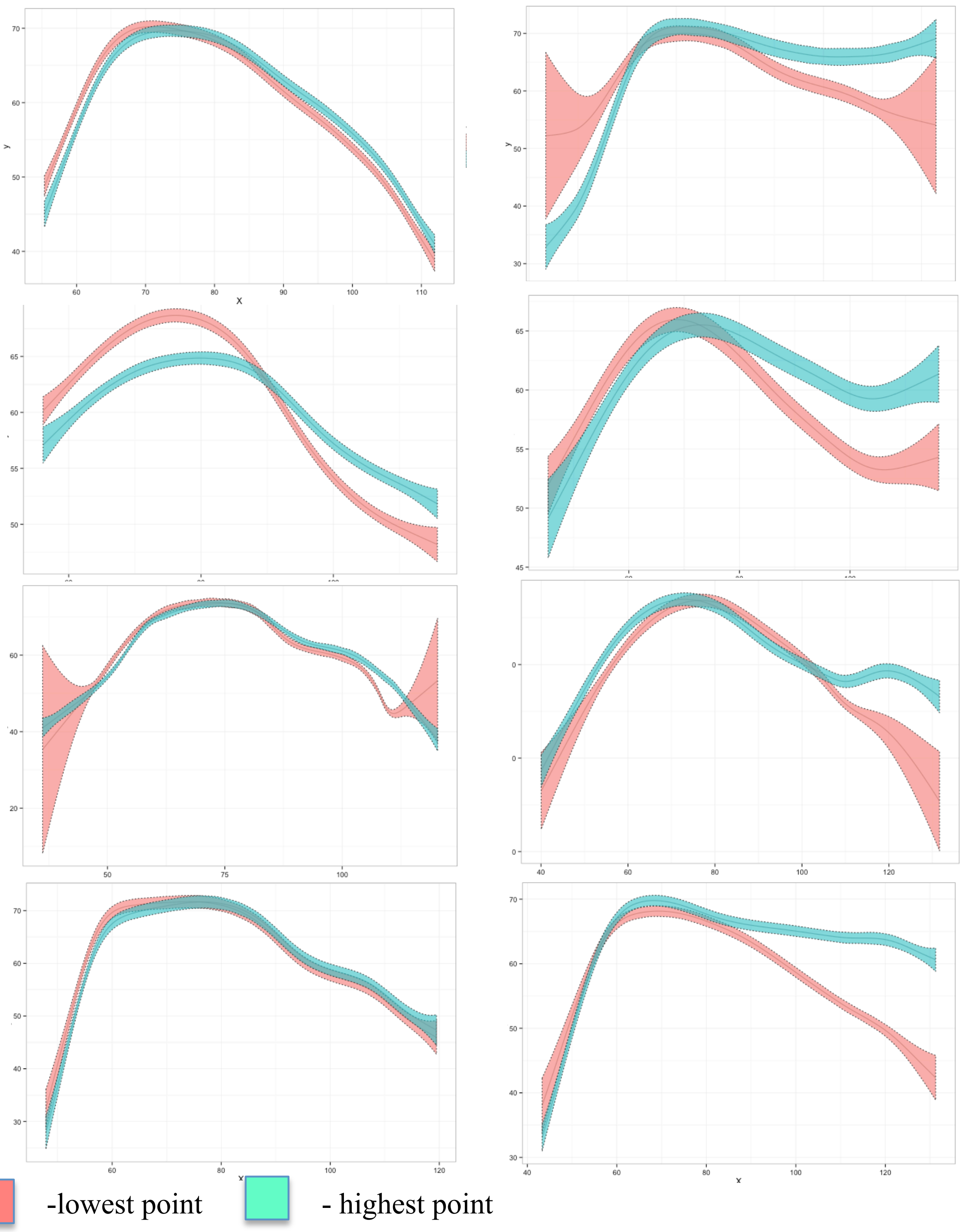
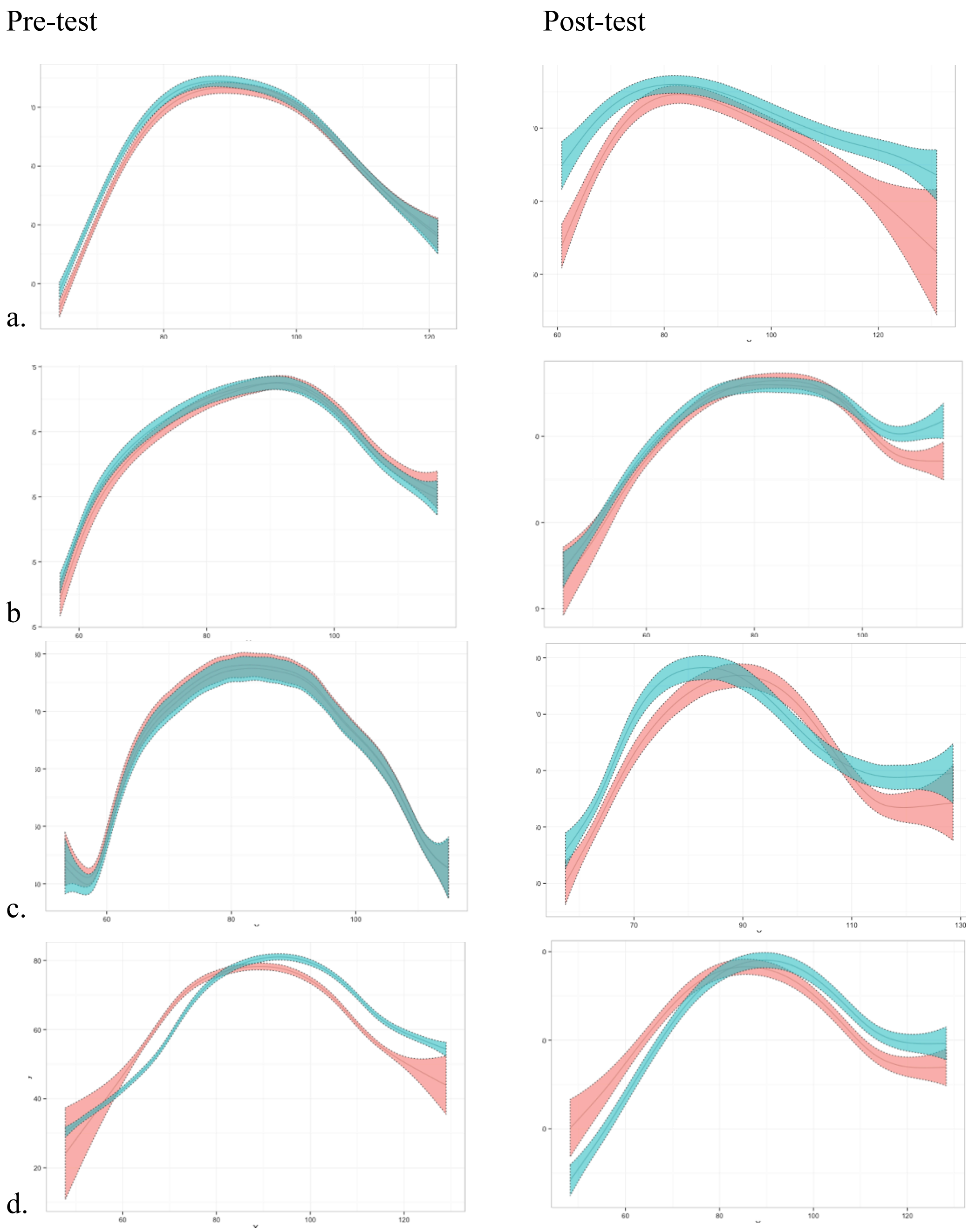


Figure 2. SS-ANOVAS for [ɔ] – ultrasound group



-lowest point - highest point

Figure 3. SS-ANOVAs for [iː] - ultrasound group

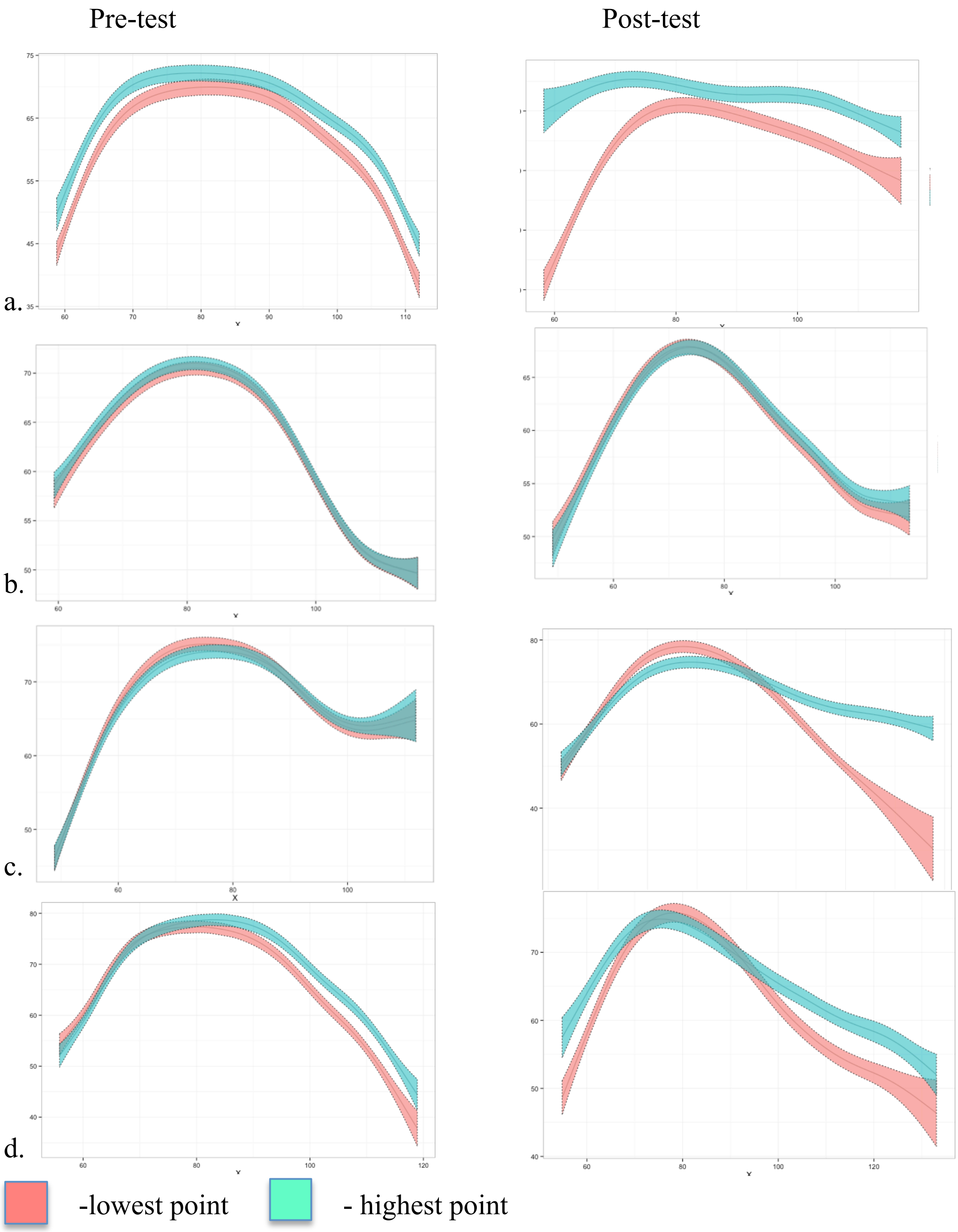


Figure 4. SS-ANOVAs for [uː] - ultrasound group

For the non-ultrasound group, the improvement does not appear to be as strong. Although in Figure 5b there is a large difference between point A and point B, the same difference is also present in the pre-test: the splines indicate that this speaker was already using tongue tip raising in the pre-test, which was corroborated in the judgments (see section 4.2). As for Figure 5a, although there is a difference between point A and point B in the post-test only, the shape of the splines is somewhat different from those in the ultrasound group. In fact, this speaker produced an [ɹ] (as per the judgments in section 4.2) in the post-test. It seems, therefore, that this speaker did raise the tongue tip after the training, but not in the correct configuration for a coda [ɫ].

The [iː] results in the non-ultrasound group strongly suggest lack of improvement after training. Three out of the 4 speakers had an overlapping point A and point B even in the post-test. The one speaker who had a separate point A and point B was b., the same speaker who was already raising the tongue tip for [ɫ] in the pre-test.

In the [uː] splines, two of the students have an overlapping point A and point B. As for the other two students, 7a was the student who produced an [ɹ] instead of an [ɫ], and 7b was the student who could do tongue tip raising in the pre-test. Her pre-tests SS-ANOVAs for [uː] do not seem to suggest tongue tip raising, but this could be due to either of the explanations from the above ultrasound section.

For most of the non-ultrasound group participants, there is no clear difference in tongue tip raising between the pre-test and post-test. Note, again, that subject b. was already able to produce [ɫ] before the training, and while she may have made some exaggerations in her pronunciation after the training, other evident differences may be due to the scanner's field of vision change between the pre-test and post-test.

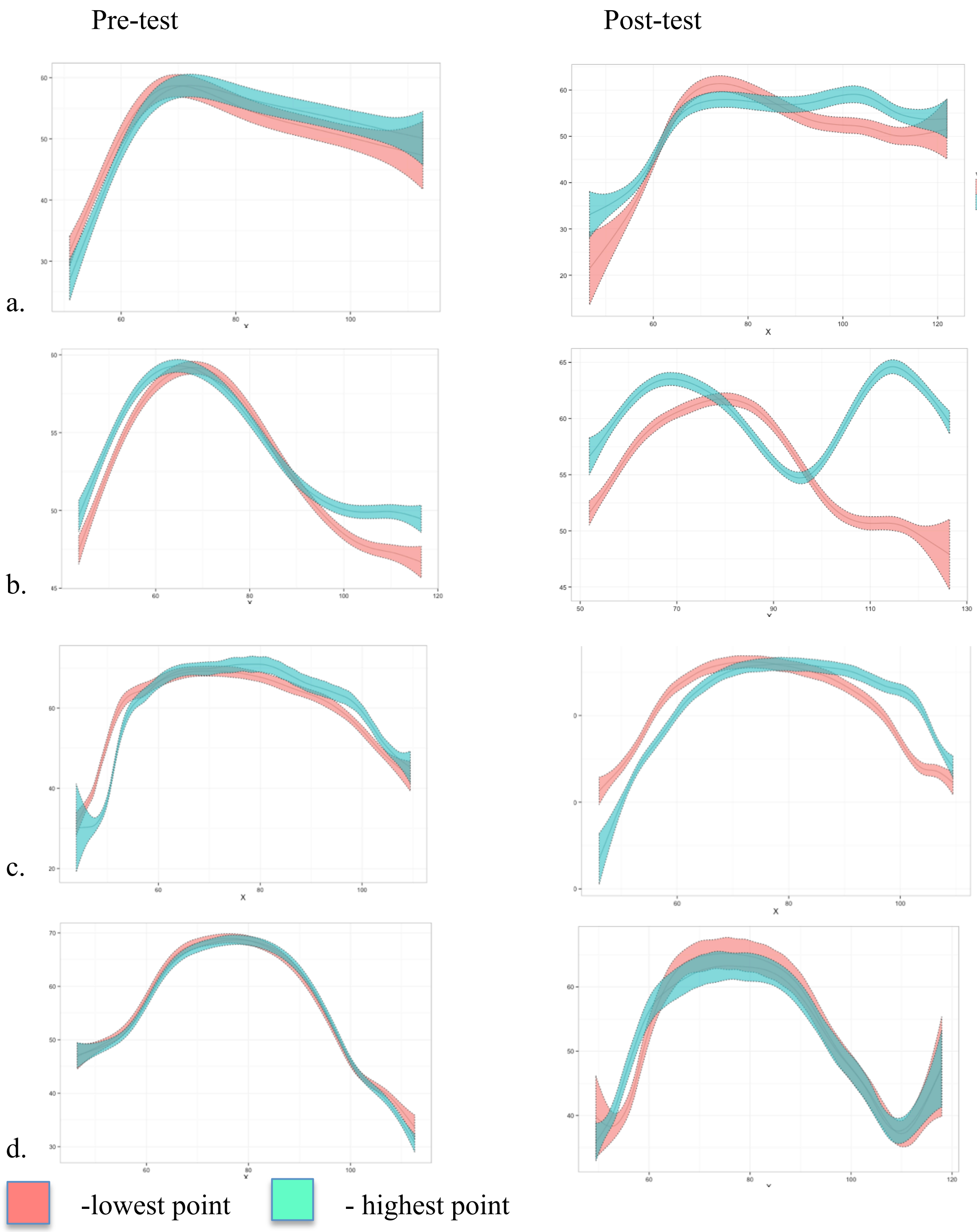


Figure 5. SS-ANOVAS for [ɔ] – non-ultrasound group

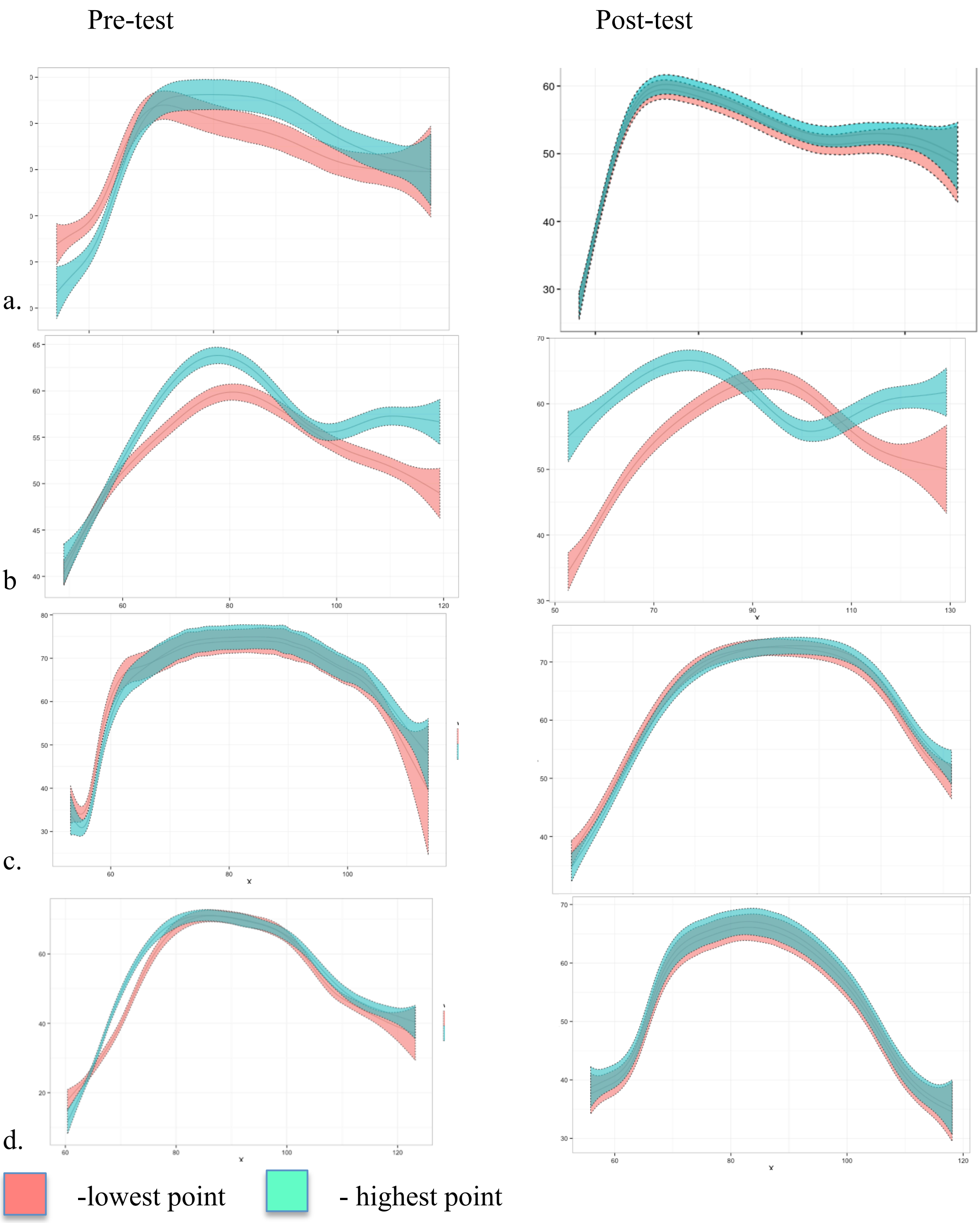


Figure 6. SS-ANOVAS for [i:] – non-ultrasound group

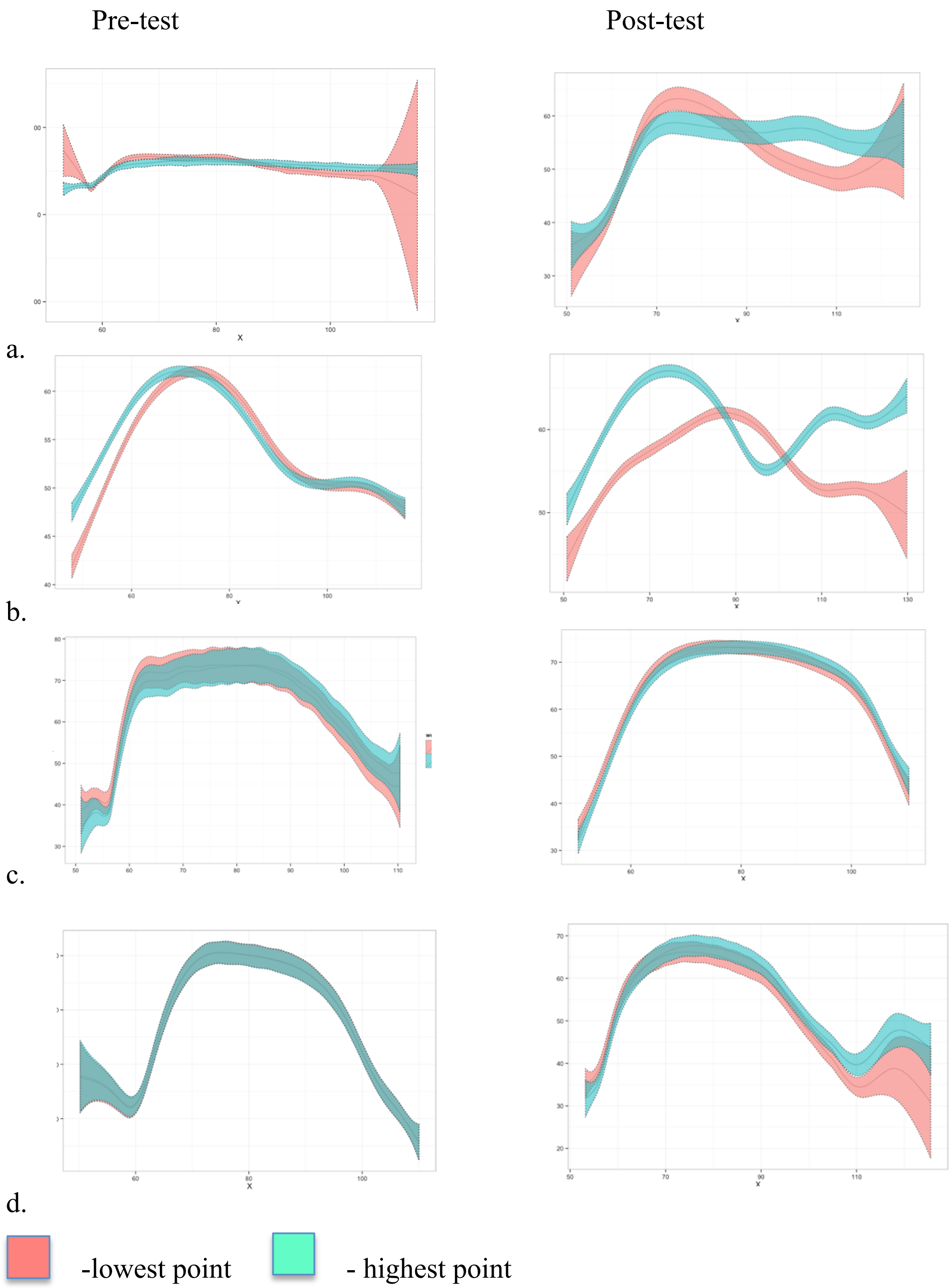


Figure 7. SS-ANOVAs for [uː] - non-ultrasound group

Judgments

To corroborate the ultrasound results, acceptability judgments were conducted by the first two authors, who gauged whether they perceived the productions to be an accurate English [ɫ]. (The judges were aware of the group identity for each of the speakers.) Each token was categorized as *[ɫ]*, *vocalized [ɫ]*, */r/*, or *not sure*. Out of 240 tokens, there were 32 disagreements between the judges (That is, one judge saying *[ɫ]* and the other saying *a vocalized [ɫ]*. There were no disagreements where one judge said *[ɫ]* and the other said */r/*.) For these 32 disagreements, a third native English-speaking judge was selected to re-judge the tokens, and her judgments were used for those tokens. For all other tokens, the judgments of the original two judges were used.

In order to conduct statistical analysis, the data were transformed into a binomial "acceptable"/"not acceptable" form, where *[ɫ]* is considered "acceptable" and *a vocalized [ɫ]*, or *an /r/* are considered "not acceptable". If one judge said *not sure*, this was ignored, and the judgment of the other judge alone was used. There were no cases where both judges said "not sure" for the same token. The data are presented below. Two one-tailed Fisher's tests (Fisher, 1954) that tested whether the improvements between the pre-test and post-test were significant were conducted separately for the ultrasound and non-ultrasound group, respectively. From the p-values, it is clear that the ultrasound group's improvement was significant, and the non-ultrasound group's improvement is not significant. These results support the ultrasound group's training being more successful than the non-ultrasound group's training.

Table 2

Results for ultrasound group's judgments

	Pre-test	Post-test
Acceptable ([ɫ])	1	49
Not Acceptable (V or r)	58	10
Total Tokens	59	59
p-value (Fisher's Test)	p< .0001 (significant)	

Table 3

Results for non-ultrasound group's judgments

	Pre-test	Post-test
Acceptable ([ɫ])	23	30
Not Acceptable (V or r)	36	29
Total Tokens	59	59
p-value (Fisher's Test)	p< .1333731 (not significant)	

DISCUSSION

Both the ultrasound spline SS-ANOVA results and the judgment results suggest that the training received by the ultrasound group may have been more effective than the training received by the non-ultrasound group. Although both groups seemed to show some improvement, the SS-ANOVAs for the ultrasound group show much clearer significant differences between the pre-test and post-test, while the non-ultrasound group's SS-ANOVA pre-test and post-test results for the most part are mostly not significantly different. The same can be said about the judgment results: the judgments of the ultrasound group indicated significant improvement, but the same judgment of the non-ultrasound group did not.

At the outset of this experiment, a high degree of [ɫ] vocalization was evident among the eight participants. This finding is noteworthy because these participants joined the study on a voluntary basis; there was no selection process that precluded students who did not vocalize [ɫ]. In other words, although this sample is extremely small, it does seem to suggest that [ɫ] vocalization is common in the L2 English of L1 Cantonese speakers. The results of the present study indicate that ultrasound may be an effective means of addressing this pronunciation difficulty.

Although the overall findings suggest improvement for the ultrasound group, there was a discrepancy between the SS-ANOVA results and the listener judgments. In the SS-ANOVA results, the improvement was most evident in the [ɔ] tokens, and, to a lesser extent, the [iː] and [uː] tokens. According to the judgments, however, the improvements were more general across the three vowel contexts. Perhaps this is because [iː] and [uː] vowels are high vowels that involve greater elevation of the tongue. Because the overall shape of the tongue is more raised, it more difficult to see raising of the tongue tip clearly on the ultrasound scans. This is partly because the ultrasound scanner cannot see the exact tip, and its position can only be extrapolated from the position of the blade. In other words, the differences in tongue shape would not be as extreme between [ɫ] and its vocalized version for [iː] and [uː]. If this is the case, then differences in these two vowel contexts may not be as evident in SS-ANOVA.

A second crucial aspect of vocalization that SS-ANOVA cannot capture is lip rounding. Regardless of tongue shape, [ɫ] will tend to sound more vocalized when the lips are rounded as in an English back vowel. This information was part of the verbal instructions to all of the students (see Appendix 1), and so neither group should have had an advantage in this regard. As for the data analysis, because the listener judgments focused on the total sound and not just the tongue shape, they were more likely to take the sound of lip rounding into account. This seems to strengthen the case for improvement of the ultrasound group across the three vowel contexts.

This investigation was a pilot study, so the amount of instruction was not ideal (approximately 30 minutes in total, but time spent with the actual ultrasound machine was even less). It seems, then, that even a small amount of ultrasound training may have affected the students' pronunciation, which bodes well for future work in this area. It must be acknowledged, however, that no subsequent tests were conducted to see whether this lesson had a lasting effect on the pronunciation of the students. The possible success of this method notwithstanding, it is unlikely that a short 30-minute lesson would have any kind of lasting effect. With this in mind, the authors are currently carrying out an expanded study that drastically increases the lessons and the amount of time that students use the ultrasound machine. A longer course of study, with multiple training sessions, would give students a better chance of retaining any articulatory adjustments they have made during the ultrasound training sessions. Additionally, the expanded study instructs students on several different segments, instead of focusing on [ɫ] solely.

By focusing on a more comprehensive range of L2 phones, this broader study will address a wider range of intelligibility issues in Hong Kong English. Foremost among these are three pairs of vowels that present difficulties to many Hong Kong students: [iː] – [ɪ]; [æ] –[ɛ]; and [uː] – [ʊ]. Many Hong Kong students find these pairs quite challenging to differentiate, which leads to decreased intelligibility. It is hoped that the use of ultrasound will allow students to better visualize distinct tongue positions for these vowels, and thereby improve their intelligibility in the process. The present, preliminary study suggests that ultrasound has the potential for success when used to address these kinds of problems.

ACKNOWLEDGMENTS

For their kind support, guidance, and recruiting abilities, the authors would like to acknowledge Principal Judy Li (李碧茵校長), and Peggy Lie (English teacher) of FDBWA Szeto Ho Secondary School, Lam Tin, Kowloon, Hong Kong S.A.R.

ABOUT THE AUTHORS

Donald White is the Native English Teacher at Queen’s College, Hong Kong, and a doctoral candidate in Linguistics at the Chinese University of Hong Kong. His principal research interest is L2 accent in a study abroad context.

Richard Gananathan is a doctoral candidate in Linguistics at the Chinese University of Hong Kong. His principal research interest is in the link between perception and production systems in human speech.

Peggy Mok is an Associate Professor at the Department of Linguistics and Modern Languages, The Chinese University of Hong Kong. Her research interests include experimental phonetics, psycholinguistics and bilingualism.

REFERENCES

Articulate Instruments Ltd. (2010). Articulate Assistant User Guide: Version 1.18. Edinburgh, UK: Articulate Instruments Ltd.

Bernhardt, B., Gick, B., Bacsfalvi, P., & Ashdown, J. (2003). Speech habilitation of hard of hearing adolescents using electropalatography and ultrasound as evaluated by trained listeners. *Clinical Linguistics & Phonetics, 17*(3), 199 - 216.

Catford, J. C. & Pisoni, D. B. (1970). Auditory versus articulatory training in exotic sounds. *Modern Language Journal*, 54, 477-481.

Chan, A. Y. W., & Li, D. C. S. (2000). English and Cantonese phonology in contrast: Explaining Cantonese ESL learners' English pronunciation problems. *Language, Culture and Curriculum*, 13(1) 337-356.

Firth, S. (1987). Developing self-correcting and self-monitoring strategies. *TESOL Talk*, 148-152.

Fisher, R. A. (1954). *Statistical Methods for Research Workers*. Oliver and Boyd, London.

Gick, B., Bernhardt, B., Bacsfalvi, P., & Wilson, I. (2008). Ultrasound imaging applications in second language acquisition. In J. Hansen & M. Zampini (Eds.), *State–of–the–Art Issues in Second Language Phonology*. Cambridge, U.K.: Cambridge University Press.

Gu, C. (2002). *Smoothing spline ANOVA models*. Springer Series in Statistics. New York: Springer-Verlag.

Tateishi, M. (2013). *Effects of the use of ultrasound in production training on the perception of English /r/ and /l/ by native Japanese speakers*. Unpublished Thesis from University of Calgary, Department of Linguistics.

Truby, H.M. (1959). Acoustic-cineradiographic analysis considerations with special reference to certain consonantal complexes. *Acta Radiologica Supplementum*, 1-227.

Tsui, H. M. L. (2012). *Ultrasound speech training for Japanese adults learning English as a second language*. Unpublished Thesis from University of British Columbia, Department of Audiology and Speech Sciences.

Appendix One – Lesson Powerpoint

Listening practice

Two gestures: first back, then forward

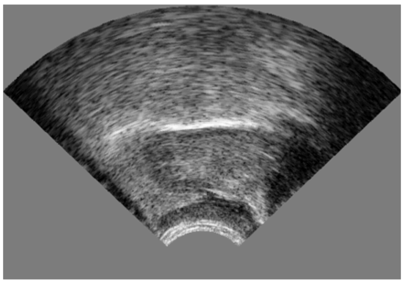
No lip rounding

Listen and watch carefully

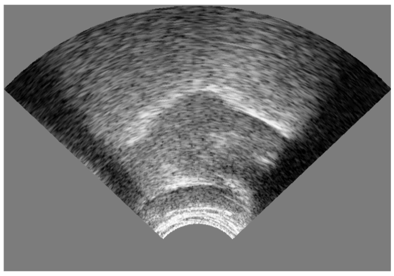
real	mall	rule
feel	fall	school
deal	hall	pool

Tongue shape differences at

Correct



Incorrect



Visualization

Let’s look at our own tongue shapes on the ultrasound.

Practice - Pairs

i	a	u
feel – few	hall - how	cool - coo
meal – mew	ball - bow	tool - two
seal – 燒	mall – 貓	fool – 膚

Practice

A – How do you feel today?
B – I’m a real mess.

A – How do the students feel after the meal?
B – A few feel sick.

A – How many students are in the hall?
B – All in all there are two hundred.

More practice

A – Where did you fall down?
B – I fell down in the hall.

A – Is the weather cool today?
B – It's too cool to go in the pool.

A – What does it say in the school rule book?
B – It says you should not act like a fool.

Contact information of authors

Donald White, The Chinese University of Hong Kong, Sha tin, Hong Kong SAR, (852) 9248 9675, donaldtimothywhite@gmail.com

Richard Gananathan, The Chinese University of Hong Kong, Sha tin, Hong Kong SAR, (852) 3943 1517, rgan@link.cuhk.edu.hk

Peggy Mok, The Chinese University of Hong Kong, Sha tin, Hong Kong SAR, (852) 3943 1768, peggymok@cuhk.edu.hk

Zhou, Z. & Li, Z. (2017). Exploring the relationship between fluency measures and speaking performance of prospective international teaching assistants. 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. 176-185). Ames, IA: Iowa State University.

EXPLORING THE RELATIONSHIP BETWEEN FLUENCY MEASURES AND SPEAKING PERFORMANCE OF PROSPECTIVE INTERNATIONAL TEACHING ASSISTANTS

Ziwei Zhou, Iowa State University, USA
Zhi Li, Paragon Testing Enterprise, Canada

Previous studies suggest that L2 fluency measures can influence the evaluation of speaking performances, but with different degrees of contribution. Such relationships are still under-researched for international teaching assistants (ITA), who play important roles in undergraduate education in the higher education institutions in North America. This study focuses on 114 prospective ITAs at a large Midwestern university in the US and aims to investigate the relationships between fluency measures and their speaking performances in an in-house speaking test for ITAs. Four categories of fluency measures, namely, speed, juncture pauses as breakdown, non-juncture pauses as breakdown, and fillers, were calculated in the form of 15 variables, based on the automated results from a modified version of Quené, Persoon, and de Jong's (2010) Praat script as well as manual annotation of the speech samples. The raw holistic scores of the speaking performance were analyzed using FACETS to obtain corrected or fair scores. A multiple regression was conducted and the results indicated that average syllable duration and normalized count of juncture pauses were the most significant predictors of the corrected score. These findings are informative for ITA programs to better understand the fluency characteristics of ITAs and their contribution to speaking performances.

INTRODUCTION

Although the notion of language fluency was used interchangeably with general language proficiency at times (e.g. "She speaks fluent French"), the impression of fluent speech was generally associated with a sense of ease, motion, fluidity, and smoothness in speech (Chambers 1997; Lennon, 2000). In speech production, fluency appears to be one of the "most easily noticeable" properties that differentiates L1 and L2 speakers (Kormos, 2006, p. 154). In his seminal work, Fillmore (1979) proposed four types of fluency in terms of language competencies: 1) the ability to fill the time with talks; 2) the ability to talk coherently with reasonable and semantically "dense" sentences; 3) the ability to speak appropriately according to specific contexts; and 4) the ability to use language creatively with novelty and imagination. Similarly, Lennon (2000) provided a working definition of fluency as "the rapid smooth, accurate, lucid, and efficient translation of thought or communicative intention into language under the temporal constraint of online-processing" (p. 26). Based on the interdependent nature of fluency issues, Segalowitz (2010) argued that fluency situated itself upon "the intersection region of the subdisciplines of cognitive science" and its inquiry should require multidimensional and multidisciplinary efforts (p. xiv).

In terms of studying fluency in L2 context, the "dual approach" is usually adopted where perceived fluency scores assigned by raters to non-native speech are compared with objective measures calculated for the same speech (Cucchiarini, Strik, & Boves, 2000; Xi & Mollaun, 2006). General findings in this line of research point to a strong positive correlation between human score and fluency measures. As reviewed by Yoon (2009), most findings revealed temporal measures of fluency to have significant roles to play in fluency scores.

More recently, researchers have paid particular attention to the possible overlapping in categorizing various fluency measures and endeavored to separate these measures by setting distinct boundaries (de Jong, Groenhout, Schoonen, & Hulstijn, 2015; Kahng, 2014). For example, de Jong et al. (2015) clustered the measures into three categories: 1) speed fluency that has been characterized as the rate and density of speech delivery; 2) breakdown fluency that concerns the extent to which a continuous speech signal is interrupted; 3) repair fluency that relates to the number of corrections and repetitions present in speech (Skehan, 2009). In particular, Kahng (2014) argued that fluency measures should be chosen in a way that they explicitly represented each aspect of fluency and should not be mathematically dependent or strongly interrelated with each other. The effort to categorize aspects of fluency and disambiguate their boundaries is necessary because one can reasonably assume that "measures from the same fluency aspect might be caused by the same cognitive problems in the speech production process" (Bosker, Pinger, Quené, Sanders, & de Jong, 2012, p. 171).

Recently in the context of language testing, there has been revitalized interest in studying fluency with more emphasis on statistical rigor, better integration of advanced speech technology, and broadened scope (e.g. including more variables such as test takers' L1 background). For instance, Ginther, Dimova, and Yang (2010) used Python interface to facilitate the automated extraction of fluency measures produced by Praat. Through administering the Oral English Proficiency Test (OEPT) via computer to 150 test takers, their analyses revealed that speech rate, speech time ratio, mean length of run, and the number and length of silent pauses were significant predictors for proficiency scores. However, fluency variables alone did not distinguish adjacent levels of the OEPT scale. Bhat, Hasegawa-Johnson, and Sproat (2010) investigated signal-level fluency quantifiers in a rated speech corpus of L2 English learners. The results of their logistic regression analyses indicated that articulation rate and phonation-time ratio significantly predicted fluency level. Using mixed-effects modeling, Bosker et al. (2012) found that listeners weighed the relative importance of the perceived fluency to arrive at holistic score or overall judgement of L2 Dutch speech. This study was further extended to 53 L2 learners of Dutch of L1 English and Turkish by de Jong et al. (2015) who validated L2 average syllable duration (ASD or inversion of articulation rate) as the most useful predictor of fluency, explaining 30% of the variance in L2 proficiency. In addition, partialling out L1 variance increased the explained variance to 41%.

Informed by the current perspective to fluency issues in language testing context, this study aims to investigate the relative contribution of fluency, as measured by its speed and breakdown domain, to the multi-componential construct of speaking proficiency (de Jong, Steinel, Florijn, Schoonen, & Hulstijn, 2012). As pointed out by Ginther et al. (2010), examining the subskills such as fluency underlying holistic score can augment our understanding in the interpretation and use of test scores and provide supporting evidence for the validity of inferences with regards to test performances. Practically speaking, information obtained from such inquiry can serve to improve fluency-related descriptors in scoring scales. Moreover, the potential usefulness of automated assessment deserves further investigations due to its advantage in providing objective measures with minimum human intervention as well as low associated expenses.

In addition, this study concerns the population of international teaching assistants (ITAs) since it has been observed that ITAs' speaking proficiency in general and speaking fluency in particular appear to be far less laudable, as compared to their high attainment of content knowledge (Gorsuch, 2011; Kaufman & Brownworth, 2006). Gorsuch 2011 particularly point out that ITAs' poor fluency, including "slow speech, false starts, and particularly pauses that violate phrasal boundaries", may pose tremendous obstacles for them to meet various academic requirements such as teaching undergraduate classes. In the light of the above accounts, this study pursues the following research questions:

1) To what extent can speed and breakdown fluency measures distinguish score levels in the ITA speaking test?
2) To what extent can speed and breakdown fluency measures predict corrected test scores in the ITA speaking test?

METHODS

The speaking test and the data. The test of interest is an English speaking test for prospective international teaching assistants (ITAs) used at a large Midwestern university in the US. The speaking test consists of two main components, namely an oral proficiency interview section and a simulated mini-lecture section which comprises a 5-minute lecturing and a 3-minute question-answering. This study only focused on the second component of the test (TEACH section) because ITAs' performance of simulated lecturing is of primary concern for individual departments and the teaching performance in the form of monologue is relatively easier to extract fluency measures, compared with the interview performances.

The speaking performance is holistically evaluated by three trained raters for overall comprehensibility, the effectiveness of oral language, and listening ability, using a 300-point score band. The scores are then converted to a 4-point scale with 4 being the lowest level (not certified) and 1 the highest level (fully certified). According to the online score guide of the speaking test, fluency is explicitly or implicitly mentioned in the descriptors of the four levels. For example, a performance at level 1 (lowest level) may be characterized as using "short utterances that are filled with hesitations, pauses, self-corrections, and ineffective reformulations" whereas a level 1 performance would "show very good fluency."

Due to the small number of level 4 test-takers, we decided not to include Level 4 samples and collected 114 TEACH speech samples with 38 sampled from each of the three levels (levels 1-3) from a pool of 227 samples rated by 11 raters. Among the 114 test-takers, 78 were males and 36 were females. The major first languages included Chinese (45), Korean (9), Vietnamese (4), Hindi (4), Nepali (4), Bengali (4), and Arabic (3). They represented graduate students from three main colleges: Engineering (58), Humanities (28), and Business (11).

Considering the fact that fluency analysis is labor-intensive and time-consuming, we limited our analysis to the first 2-minute segment of each teaching monologue sample, excluding the period of silence at the beginning of the recording as well as long non-speech pauses for blackboard writing, if present. These segments were pre-processed with Audacity 2.0.5 to normalize amplitude level and to remove background noise.

Fluency measures. Both automated measures and manual annotation-based measures of fluency were employed in this study. A Praat script written by Quené, Persoon, and de Jong (2010) was then used on Praat 6.0.17 to analyze the 2-minute segments for three automated measures of the speed aspect of fluency, i.e., speech rate (number of syllable/total time), articulation rate (number of syllables/phonation time), as well as average syllable duration (ASD, phonation time/number of syllables).

The resultant textgrid files from Praat output were used as a basis for manual annotation of breakdowns with AS-unit and clausal unit (Foster, Tonkyn, & Wigglesworth, 2010) as units of analysis. Three types of breakdown were annotated: juncture pause, non-juncture pause, and fillers. The juncture pauses are the noticeable silences (whose duration is greater than 200 ms) occurring at the boundaries of clausal units as in "we are going to learn this [pause]" or sub-clausal units that can be elaborated into complete semantic unit as in "Ok. [pause] Now, let's move to …". The non-juncture pauses are the noticeable silences occurring within clausal units as in "supply chain is [pause] managed by …". The fillers include non-nasalization fillers like "uh" and "eh" and nasalization fillers like "um" and "un" either occurring separately or attached as an elongated vowel coda.

The manual annotation was carried out by the two researchers. After initial familiarization of the annotation scheme and calibration of annotation on a set of five speech samples, each researcher annotated another set of five speech segments. 10% of the samples were later double annotated to examine inter-coder reliability. The agreement for pauses was 87% and that for fillers was 72%. The disagreement was solved through discussion which prompted a round of self-check of the annotations to improve accuracy.

The combination of automated measures and manual annotation contributed a total of 15 fluency measures to reflect four major aspects of fluency, namely, speed, juncture pauses as breakdown, non-juncture pauses as breakdown, and fillers (see Table 1 in the Results section). As mentioned earlier, the speed aspect was represented with speech rate, articulation rate, and ASD. The breakdowns were characterized with count, duration, as well as mean length of two types of pauses. In addition to these variables, we conceptualized density of speech as another aspect of fluency in terms of mean length of run (total length of sounding segment/number of sounding segment), count-based ratio (ratio of count of sounding segment by count of pauses), and duration-based ratio (the ratio of duration of sounding segment by duration of pauses).

To make the variable values comparable across speech samples, normalizations are applied to count and duration variables separately. This is different from previous practice (e.g. Kahng, 2014) where both variables are normalized against total speaking time. Thus, For example, the normalized count of juncture pause is the result of number of juncture pause count divided by the number of sounding segments given by Praat script output. Likewise, the duration of juncture pause is normalized against the corresponding phonation time.

Speaking proficiency. In this study the speaking proficiency is treated as the dependent variable and multiple fluency measures are the independent variables or predictors of speaking proficiency. Like other performance assessment, raters played an important role in the speaking test through their operationalization of the rating scale and subjective evaluation of test-takers' performance, which may introduce variability of rater severity to the final ratings. To account for this potential impact, we used multifaceted Rasch model (MFRM) to estimate rater severity with the complete date set (N=227), which was then utilized to adjust the raw ordinal scores and produce "fair scores" of interval nature on the same reporting scale (Linacre, 2009). The multifaceted Rasch model, as an extension of the Rasch model, is capable of calibrating situational factors such as rater, testing occasions, task formats, along with the traditional parameters like test-taker ability, item difficulty using a common interval scale in the unit of logit.

Data analysis. In compliance with the variable distribution requirement from Ordinary Least Square method (OLS), we started with an examination of the normality assumption for each of the independent variable. Logarithmic or square root transformation was applied to the non-normal variables (see Table 1). In addition, a series of analysis of variance (ANOVA) were conducted to compare the independent variables across proficiency levels, followed with pairwise comparisons using Tukey HSD method.

The relationship between the fair scores and the independent variables was modeled using stepwise multiple regression. The assumption of multicollinearity was examined using the correlation matrix of the independent variables and the variance inflation factor (VIF) values of each independent variable in regression models. The assumptions of residual normality, linearity, and residual variance were checked through visually examining the corresponding scatterplots. The final model was determined based on the comparison of the Akaike's Information Criterion (AIC) values and the adjusted R^2 values of the nested models as well as the theoretical soundness of the models.

RESULTS

This section reports the results of FACETS analysis, pairwise comparisons as a part of ANOVA procedures, as well as stepwise multiple regression to answer the research questions.

FACETS analysis. The mean fit statistics for the rater facet and test-taker facet were examined. The mean infit and outfit mean square values of test-taker facet were 0.79 and 0.77, respectively, whereas the mean values for the rater facet were 0.88 and 0.69. Overall, the mean fit statistics were in the range of 0.5 and 1.5 for acceptable model fit, while 1 rater and 63 test-takers (out of 227) were identified as misfit. The chi-square test value of the rater facet was statistically significant ($\chi^2 = 30.9$, $df = 9$, $p < .01$) and the rater facet did exhibit variation in rater severity with a range -2.18 to 1.17 logits and a mean of zero, which warranted the use of adjusted scores or fair scores.

Descriptive statistics and ANOVA. Table 1 contains the means and standard deviations of the fair scores and the independent variables. To save space, an additional column was inserted to the right of Table 1 to report the statistically significant pairs based on the results of pairwise comparisons from the ANOVA procedures.

As expected, the significant differences in fair scores existed across the three proficiency levels. By contrast, five out of 15 independent variables did not show significant differences across the proficiency levels: duration of juncture pauses, mean length of juncture pauses, mean length of non-juncture pauses, mean length of fillers, and mean length of run. Three independent variables showed significant differences between levels 1 and 3 only (number of non-juncture pauses, duration of non-juncture pauses, and count of fillers). In addition to the identified difference between levels 1 and 3, three were also different between levels 2 and 3 (duration of fillers, count-based ratio, duration-based ratio), and four variables were also different between levels 1 and 2 (speech rate, articulation rate, ASD, and number of juncture pauses).

The pairwise comparison results indicate that some independent variables may be more effective in predicting speaking proficiency in this study while the utility of others may be questionable such as the mean length-based variables.

Table 1

Descriptive statistics by proficiency levels and results of pairwise comparisons

	Level 1	**Level 2**	**Level 3**	**Pairwise comparison** [a]
fair score	1.11 (0.08)	1.99 (0.35)	2.81 (0.35)	1-2, 1-3, 2-3
speech rate	3.09 (0.46)	2.84 (0.57)	2.68 (0.38)	1-2, 1-3
articulation rate (log) [b]	1.39 (0.15)	1.30 (0.17)	1.26 (0.10)	1-2, 1-3
ASD (log)	-1.39 (0.15)	-1.30 (0.17)	-1.26 (0.10)	1-2, 1-3
count of juncture pause	0.48 (0.10)	0.42 (0.10)	0.38 (0.08)	1-2, 1-3
duration of juncture pause (log)	-1.82 (0.33)	-1.86 (0.45)	-1.80 (0.36)	none
mean length of juncture pause (log)	-0.53 (0.23)	-0.48 (0.26)	-0.39 (0.28)	none
count of non-juncture pause	0.49 (0.10)	0.54 (0.14)	0.58 (0.13)	1-3
duration of non-juncture pause (log)	-2.04 (0.38)	-1.90 (0.71)	-1.68 (0.50)	1-3
mean length of non-juncture pause	0.47 (0.09)	0.48 (0.10)	0.52 (0.10)	none
count of fillers (sqrt)	0.38 (0.17)	0.43 (0.17)	0.52 (0.17)	1-3
duration of fillers (sqrt)	0.19 (0.09)	0.21 (0.09)	0.27 (0.10)	1-3, 2-3
mean length of fillers (log)	-0.90 (0.22)	-0.99 (0.17)	-0.90 (0.17)	none
mean length of run (log)	0.52 (0.19)	0.48 (0.30)	0.43 (0.19)	none
count-based ratio	0.89 (0.10)	0.87 (0.10)	0.80 (0.10)	1-3, 2-3
duration-based ratio (log)	1.08 (0.29)	0.99 (0.46)	0.81 (0.32)	1-3, 2-3

Note: a. The pairwise comparison column shows statistically significant pairs only. b. log = logarithmic transformation, sqrt = square root transformation

Multiple regression. The correlation matrix of the independent variables indicates two pairs of highly correlated variables (over .9): articulation rate and ASD, duration of non-juncture pause and duration-based ratio. To avoid including the variables measuring similar constructs, we removed articulation rate and duration-based ratio, which left 13 variables for multiple regression analysis.

A final model was determined based on the comparison of AIC values and adjusted R^2 values of the nested models. Among the remaining 13 variables, only five variables were retained in the final model with two being significant predictors, i.e., ASD and number of juncture pause (see Table 2). There were no multicollinearity issues associated with the variables. The multiple R^2 value of the final model was .20, which means that about 20% of the fair score variance can be explained with the final model.

Table 2.

Predictor variables in the final model

Coefficients:	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**
(Intercept)	5.024	0.732	6.866	<.001 ***
ASD (log)	0.947	0.476	1.988	.049 *
count of juncture pause	-1.750	0.866	-2.022	.046 *
mean length of juncture pause (log)	0.392	0.313	1.252	.213
duration of non-juncture pause	-0.112	0.172	-0.649	.518
count-based ratio	-1.269	0.692	-1.834	.070

Residual SE: 0.6948 (103), Multiple R^2: 0.2012, Adjusted R^2: 0.1625, F-statistic: 5.19 (5, 103), $p < .001$

The regression coefficient of the log-transformed ASD (average syllable duration) was 0.947. In other words, a 10% increase in the log-transformed ASD will bring about an increase of 0.039 of fair scores (0.9471*log(1.1) = 0.039) or to lower the speaking proficiency level (level 1 is the highest and level 4 is the lowest level in the test) with other variables being held constant. The regression coefficient of the normalized number of juncture pauses was -1.75. This suggests that with an increase of 0.1 unit of normalized number of juncture pauses (number of juncture pause/number of sounding segments), the fair score will decrease by 0.175 points with other variables remaining unchanged. In the final model, the variable count-based ratio had a close-to-significant-level p-value (0.070) and its regression coefficient was -1.26. Similar to the variable normalized number of juncture pauses, this density variable had a negative impact on the fair score.

CONCLUSIONS AND DISCUSSIONS

Our findings about the relationships between fluency measures and speaking proficiency are, to some extent, in line with those from previous studies. In this study the importance of average syllable duration was shown in its role as a significant predictor as well as its capability of distinguishing proficiency levels 1 and 3. Similar contribution of the speed measure was also reported in de Jong et al. (2015). Other studies have highlighted pauses as important indicator of fluency, but very few made distinction between different types of pauses as this study did. Our findings about the count of juncture pauses as significant predictor confirmed that proper pauses like pauses at the boundary of grammatical units can be positively perceived. It should be noted that the explanatory power of the final model is limited partly because fluency is only one aspect of the speaking construct as embodied by the proficiency levels. Nevertheless, the findings can be used to inform both rater training for the speaking test as well as ITA instructions so that fluency can be better operationalized and taught.

Methodologically, this study covered a variety of fluency measures yielded from a combined approach with automated analyses and manual coding. Of course, there are other measures which could help predict speaking proficiency levels, for example, repair as a breakdown measure and pauses at phrasal boundaries. Future studies should take them into account as well. In addition, different modeling methods should be tried, including automated speech recognition (ASR)-based acoustic modeling and corpus-based language modeling.

ABOUT THE AUTHORS

Ziwei Zhou is a PhD student in Applied Linguistics and Technology from Iowa State University. He holds a Master's degree in TESOL from the University of Pennsylvania. His research interests include automated assessment of speaking proficiency, pronunciation learning and testing, speech processing, and natural language processing. He has presented his work in Pronunciation for Second Language and Learning conference and Midwestern Association of Language Tester conference.

Zhi Li is a Language Assessment Specialist at Paragon Testing Enterprises, BC, Canada. He holds a PhD degree in the applied linguistics and technology from Iowa State University. His research interests include language assessment, computer-assisted language learning, corpus linguistics, and systemic functional linguistics. He has presented his work at a number of professional conferences such as AAAL, LTRC, and TESOL. His research papers have been published in *System* and *Language Learning &Technology*.

REFERENCES

Bhat, S., Hasegawa-Johnson, M., & Sproat, R. (2010). Automatic fluency assessment by signal-level measurement of spontaneous speech. In *2010 INTERSPEECH Satellite Workshop on Second Language Studies: Acquisition, Learning, Education and Technology*, Makuhari, Japan.

Bosker, H. R., Pinget, A.-F., Quene, H., Sanders, T., & de Jong, N. H. (2012). What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing*, *30*(2), 159–175.

Chambers, F. (1997). What do we mean by fluency? *System*, *25*(4), 535–544.

Cucchiarini, C., Strik, H., & Boves, L. (2000). Quantitative assessment of second language learners' fluency by means of automatic speech recognition technology. *The Journal of the Acoustical Society of America*, *107*(2), 989–99.

de Jong, N. H., Steinel, M. P., Florijn, A. F., Schoonen, R., & Hulstijn, J. H. (2012). Facets of speaking proficiency. *Studies in Second Language Acquisition*, *34*(1), 5–34.

de Jong, N. H., Groenhout, R., Schoonen, R., & Hulstijn, J. H. (2015). Second language fluency: Speaking style or proficiency? Correcting measures of second language fluency for first language behavior. *Applied Psycholinguistics*, *36*(2), 223–243.

Fillmore, L. W. (1979). Individual differences in second language acquisition. In C. Fillmore, D. Kempler, & W. Y. S. Wang (Eds.), *Individual differences in language ability and language behavior.* New York, NY: Academic Press.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, *21*(3), 354–375.

Ginther, A., Dimova, S., & Yang, R. (2010). Conceptual and empirical relationships between temporal measures of fluency and oral English proficiency with implications for automated scoring. *Language Testing*, *27*(3), 379–399.

Gorsuch, G. J. (2011). Improving speaking fluency for international teaching assistants by increasing input. *TESL-EJ*, *6*(4), 1–25.

Kahng, J. (2014). Exploring utterance and cognitive fluency of L1 and L2 English speakers: Temporal measures and stimulated recall. *Language Learning*, *64*(4), 809–854.

Kaufman, D., & Brownworth, B. (2006). *Professional development of international teaching assistants*. Alexandria, VA: TESOL.

Kormos, J. (2006). *Speech production and second language acquisition*. Mahwah, NJ: Lawrence Erlbaum.

Lennon, P. (2000). The lexical element in spoken second language fluency. In H. Riggenbach. (Ed.), *Perspectives on fluency* (pp. 25-42). Ann Arbor, MI: University of Michigan Press.

Linacre, J. M. (2009). *FACETS Rasch-model computer program*. Chicago, IL: Winsteps.com.

Quené, H., Persoon, I., & de Jong, N. (2010). *Praat Script Syllable Nuclei v2*. Retrieved from https://sites.google.com/site/speechrate/Home/praat-script-syllable-nuclei-v2

Segalowitz, N. (2010). *Cognitive bases of second language fluency*. New York, NY: Routledge.

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, *30*(4), 510–532.

Xi, X., & Mollaun, P. (2006). Investigating the utility of analytic scoring for the TOEFL Academic Speaking Test (TAST). *ETS Research Report Series,* 2006(1), i-71.

Yoon, S.-Y. (2009). *Automated assessment of speech fluency for L2 English learners.* Unpublished doctoral dissertation. University of Illinois at Urbana Champaign.

Bowman, D. (2017). Oye me canto, mi son: Using tongue twisters and songs. 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. 186-190). Ames, IA: Iowa State University.

TEACHING TIP

OYE MI CANTO, MI SON
USING TONGUE TWISTERS AND SONGS

Douglas Bowman, American Association of Teachers of Spanish and Portuguese

This teaching tip was developed to be used with middle and high school students at the beginning of week 2 in level I Spanish. The practice was for all learners within the classroom to practice at the same time. Class sizes ranged from 27 to 35. The tip was first implemented in 2000 and modified as students improved in performance and increased their willingness to participate. Because of its success, it was also applied at levels II and III with students who were not exposed to the tip in level I. These students also reacted with a strong willingness to participate. Starting in 2008 the principles of the *30 Million Word Initiative* (Suskind, 2015) were added: high repetition in a meaningful way with negotiated meaning. From this sprang the concept for the *1 Million Word Initiative* (the author's initiative). This initiative has the goal that by the end of the 3rd year of high school exposure, students should have the opportunity to hear and react to 1 million words with repetition in their L2 Spanish experience.

The approach began as an attempt to reconcile 2 different writing conventions, English and Spanish, that use the same symbols. Sound-symbol correspondences differ in both languages. Spanish has a more transparent correspondence, while English has a more opaque one. This particular teaching tip applies to the letter <*o*> in Spanish, and is for L1 English speakers.

The tip recognizes that connecting sound and spelling needs experiential practice to modulate the vocal chords to tune, reproduce and differentiate the [o]. Minimizing the length of the text and high repetition within the text are critical to making this connection automatic. The repetition of a specific sound also allows for frequent self-evaluation by the learner. The initial use of the sample is for sound-symbol correlation. But repetition continues into discussion and other practice, encouraging more meaningful repetition. For example, a simple translation should be orally provided. Kinesthetic practice can be employed and a discussion can follow. As learners become more comfortable with the text, they may be able to negotiate meaning. Other themes presented by the selections may also be brought up for discussion as the selection is used again in subsequent classes.

There are several ways to practice sound-symbol correspondences for L2 Spanish. The chosen series for the teaching tip is specific to the pronunciation of the <o>. All practices follow the same format: Select a tongue twister or song with multiple examples of the sound to be practiced.

1. Use of kinesthetic movement to encourage listening
2. Discussion of the text
3. Repetition of the text several times in subsequent days in its full context.

A key component for my success with this technique is the high number of repetitions of a particular sound in rich context. This is a process that occurs in within 70 to 100 milliseconds (Christiansen, 2016) when the brain begins to initiate longer term memory.

Two different exercises are provided to fulfill the goals of the teaching tip. To reiterate, there are 3 goals for this specific teaching tip.:

1. To demonstrate examples and practice of the sound-symbol correspondence for <*o*>
2. To provide specific samples for the <*o*> as the final sound of a word, a difficult context for L1 English speakers learning Spanish
3. To provide samples the new sound-symbol correlation in a manner compatible with ideas of comprehensible input to begin automatization of pronunciation skills.

Exercise 1: This uses the idea of strip stories to help learners engage in listening to the L2. Choose a tongue twister or song with the sound-symbol correspondence to be practiced.

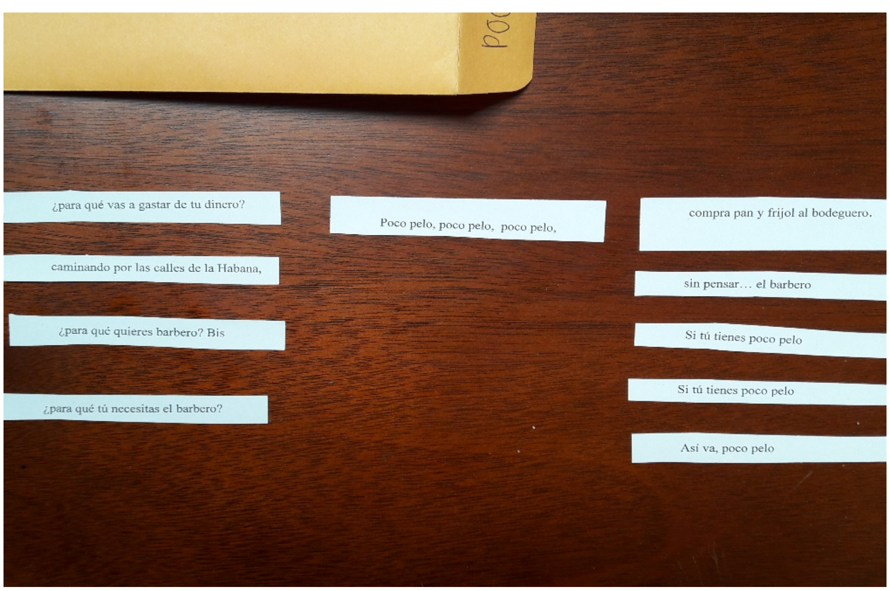
Preparation: Select the lyrics for the song or tongue twister and check for conventional spelling and other symbols such as commas, etc. In accessing lyrics on-line, care needs to be taken because words are often misspelled for various reasons.

Double space the lines of the text.

1. Print the lyrics, choosing paper of different colors. Quite often lines will end up on the floor, different colors aid in retrieving them.
2. Cut the lines without giving obvious hints as to how they go together. The photograph gives an example of how this is done.
3. Provide one set to each learner.
4. When finished with one class, keep each set in separate envelopes so that the exercise is ready for the next group of learners.

Implementing the strategy:

1. When the learner receives the envelope and places the lines on the table, suggest that half the lines go to the left and half to the right with enough space in-between to move the lines as they are identified.



2. Instructor reads the tongue twister at normal speed or plays the song all the way from beginning to end without stopping.
3. Encourage learners to place any identified line in the middle. The other lines will be placed above or below as they are identified in subsequent reading of tongue twister or playing the song. In the beginning, this process needs from 5 to 8 repetitions before learners have sequenced the lines. As the year progresses, students may need fewer repetitions to sequence the lines.
4. When just a few learners need to finish, place a correct version of the item in question on the board and have learners compare with the correct version.
5. Discuss any words with unclear meanings or pronunciation.
6. Provide a short translation.
7. Discuss the selected reading.
8. Have students place the paper strips in their envelopes and collect them.
9. From the version on the board, repeat the text chorally.
10. On following days, only step 9 needs to be performed.

Paco Peco	(Boy’s nickname)
chico rico	rich child
le gritaba	yelled
como loco	crazyly
a su tío	at his uncle
Federico,	Frederick,
y este dijo,	and this one said,
no por poco,	without exaggerating, (not by much)
Paco Peco	(boy’s nickname)
poco pico.	says little (little beak)

With repeated practice, some learners may need to know that:

1. Like English, Spanish uses upper case letters in proper names and the initial word of a sentence or title.
2. Each line has 4 syllables.
3. An accent mark identifies 2 separate syllables, not a blending or a diphthong.
4. According to Spanish pronunciation conventions, line 7 also has 4 syllables. (When a word ends in a vowel and the next word begins with a vowel, the vowel is pronounced as a diphthong.)
5. Pronouncing with 4 syllables per line helps with producing the tongue twister more quickly.

The assessment

The learner reads aloud the tongue twister twice, once at normal speed for accurate sound production, and the second as fast as possible, as any tongue twister. Though there are 22 words, by the time the learner finishes the assessment, the odds are that these words have been repeated more than 10 times each. I use a grading scale of 90% for the letter <o> and 10% for the rest of the tongue twister, combined for both attempts.

Exercise 2: To help the learner improve the new sound-symbol correspondence a new exercise is introduced after the tongue twister and before the assessment.

Repeat the different stages described in Exercise 1 above. I use the Orquesta Aragón version of *Poco pelo*. This helps the students repeat the <o> words in a new, yet largely familiar context. It also opens the class to greater discussion of grammar, culture, and contextual meaning. For more information, please contact the author.

Poco pelo, poco pelo, poco pelo, ¿para qué necesitas el barbero? (bis) (repeat)	Little bit of hair(x3) why do you need the barber?
¿Si tú tienes poco pelo, para qué vas a gastar de tu dinero? ¿Si tú tienes poco pelo, para qué tú necesitas el barbero?	If you have little hair, Why are you going to spend your money? If you have Little hair, Why do you need the barber?
Así va… poco pelo caminando por las calles de la Habana, sin pensar… el barbero, compra pan y frijol al bodeguero.	So he goes… little hair wandering the streets of Havana without thinking… the barber. Buys bread and beans from the shopkeeper

On a subsequent day, when the song is presented the fifth time, or later, a variation in its performance induce the learners to re-engage with the pronunciation and content. For example, the class could read the lyrics aloud as a chorus and as an orchestra, with the instructor identifying by hand the different group of learners to read aloud as subgroups, paying attention to where the artist is in the song. An understanding of sound-symbol correlation is reinforced by staying synchronized with the artist as the song is performed. As can be seen, the sound-spelling practice with <o> is extremely rich in usable information that furthers access to the L2 culture and therefore to communication.

ABOUT THE AUTHOR

Douglas Bowman is a retired Spanish teacher with 30 years of experience at Head Start and secondary levels. He has tutored in Japanese and German and is the editor of linguisticsnapshot.org. He is interested in the maintenance of endangered languages. He has presented internationally in Canada, Puerto Rico, Costa Rica and Panama. He has presented at state organizations in Colorado, Illinois, Iowa, Kentucky, Michigan, Oregon and Virginia, as well as Central States Conference on the Teaching of Foreign Languages (CSC) and AATSP national in New Mexico and Florida.

REFERENCES

Christiansen, M. (2016, June 10). 'Now-or-never bottleneck' explains language acquisition. *ScienceDaily*. Retrieved November 6, 2016 from www.sciencedaily.com/releases/2016/06/160610140815.htm

Suskind, D. (2015). *30 million words, Building a child's brain*. Penguin.

Chan, M.J. (2017). Anytune slows down sound tracks for language 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. 191-194). Ames, IA: Iowa State University.

TEACHING TIP

ANYTUNE SLOWS DOWN SOUND TRACKS FOR LANGUAGE PRACTICE

Marsha J. Chan, Sunburst Media and Mission College, CA

INTRODUCTION

Anytune™, a slow-downer app designed for musicians and singers, allows you to adjust the tempo of a music track without changing the pitch. Learning to speak a new language fluently is like learning music. When a phrase, dialog, story, or speech is played through Anytune, a learner can slow it down and practice it easily. Anytune loads all songs in your iTunes library; then you pick a track to practice. Within the track, you can set A and B points around difficult phrases to create a loop that plays at a percentage of the original tempo. The loop automatically restarts and plays from A to B so that you can practice the phrase again and again. The Step-it-up Trainer function repeats a section, incrementally increasing the speed from 50% to 100% in ten repetitions. The tempo and number of repetitions can be adjusted to your liking. These features allow you, as a teacher, to tailor the way you present a recorded model to your students. Students using Anytune can use these controls independently to build pronunciation accuracy, speed, rhythm, expression, and fluency. I have no connection with Anytune developers, but I do find this 5-star music practice app useful for personal use as well as for my language learning students. This article will explain how to use a selection of its features that are especially useful for language learners.

Download and install the app

Anytune[1] is available for Mac OS and iOS devices (iPad/iPod/iPhone); purchase it from Apple iTunes App Store. I use the free version of Anytune for iPhone and iPad; a student of mine prefers Anytune for Mac, which is available for purchase from the Mac App Store. From the free app, you can make in-app purchases for additional features. At the time of this writing, the Anytunes.us website states that versions for Windows OS and Android are coming.

Import sound files

First, make sure that all of your desired language sound files (AKA "songs") have been imported into the iTunes library on your Mac or Windows computer or iCloud. You may want to put these into a playlist separate from your real music.

Second, sync your iPad[2] to your iTunes library, making sure that your new playlist, if you created one, is check-marked. Please note: the audio files must be synced from iTunes; if you create a file on your iPad or download one from the Internet, you must upload it to your iTunes library and then sync it to the Music library on your iPad.

---

[1] Anytune, Inc. http://anytune.us
[2] I will herein use iPad to refer to iPhone and iPod as well

Third, import the desired file(s) into the Anytune library from the Music library on your iPad. When you launch Anytune, you'll see a message: "Anytune would like to access Apple Music and your media library." Click OK to allow this. Then "Choose a song"–you'll be able to choose from your whole Music library. To find a song, tap the "<" on iPhone or "Songs" on iPad. Tap the track to import it. You can browse content in your Music library by Songs, Playlists, Artists, and Albums. (Reminder: the "song" audio file must reside in the Music library on your device.)

Play a sound file

When you choose a sound file, Anytune first analyzes it. In the major part of the main view, you will see the sound formation, which can be zoomed by pinching, and a red vertical line designating the position of the playhead.

Tap the triangular ▸ at the bottom of the screen to play the sound; it will play from the beginning to end until and unless you press || to pause it.

Buttons and menu items will glow blue when activated or selected.

Control the tempo

Slowing down the model helps learners practice the pronunciation of sounds as well as the phrasing, rhythm, intonation, and expression. The Tempo control is on the left side of the screen. Tap the – (minus) to slow down the sound tap the + (plus) to speed up the tempo. Changing the tempo in Anytune does not change the pitch. (If you need to change the pitch for music, tap the flat and sharp symbols.) Experiment with different starting tempos–say, 50% or 65%–in addition to the number of attempts at a given tempo before speeding it up.

Create a loop and practice a segment repeatedly

Often it's better to practice short segments of speech, especially ones that are more challenging for the speaker. This is where a loop comes in handy. Listen and find the segment that you'd like (your students) to practice. Press the oval loop between the Tempo control and Pitch control to reveal the Loop controls. Tapping it again returns it to the Tempo controls.

Tap **A** to designate the start of a loop. Tap **B** to designate the end of a loop.

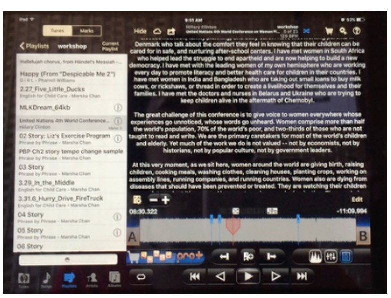
To enable Looping, tap the oval Loop in the bottom left corner. The looped section turns orange.

Once you've created a loop, you can change the practice segment by dragging the **A** and **B** with your finger to different start and stop spots.

Step it up from slower to normal tempo

Besides controlling the tempo manually, you can have Anytune start slow and step up the tempo each time it plays. Tap the icon that looks like steps with a loop. The Step-it-up Trainer plays your sound file, in its entirety or in a looped segment, at automatically increasing speeds. The default is 50% to 100% in ten steps. Unless it's a manageable utterance, I recommend using the Step-it-up Trainer after a student has practiced difficult segments in loops prior to attempting the entire sentence, paragraph, or longer speech.

Transcript



Instead of reading the text (transcript) of your sound file on paper or in a book, you can read it on the iPad while playing the sound file in Anytune. Tap the Mode selection near the bottom right (the one that looks like text opens the Lyrics window). You can enter the text manually on the iPad in Anytune or prior to syncing from iTunes to your device's Music library. By the way, popular song lyrics and transcripts of public speeches such as Hillary Clinton's plenary speech to the United Nations 4th world Conference on Women "Human rights are women's rights and women's rights are human rights" shown in this figure are readily available on the web–you can copy and paste them in the Lyrics section of the sound file.

Support

You can watch the video clip I created on this topic here: https://youtu.be/Rx-K3ydYYFk
The developer provides a user guide on its website. Anytune iOS User Guide
http://anytune.us/ios-user-guide/
Video tutorials are available on the Anytune Youtube Channel
https://www.youtube.com/user/AnytuneApp

ABOUT THE AUTHOR

Marsha J. Chan is an educational consultant, speaker, owner of a small business, Sunburst Media for Language Learners, and Professor Emerita of English as a Second Language at Mission College, Santa Clara, California. She has presented over 250 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 2000 free instructional videos at http://www.youtube.com/PronunciationDoctor.

Imber, B., Maynard, C., Parker, M. (2017) Using Praat to increase intelligibility through visual feedback. 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. 195-213). Ames, IA: Iowa State University.

TEACHING TIP

USING PRAAT TO INCREASE INTELLIGIBILITY THROUGH VISUAL FEEDBACK

Brenda Imber, PhD, University of Michigan
Carson Maynard, MA, University of Michigan
Maria Parker, MS, Duke University

In our oral skills courses, we use various feedback tools to help students learn how to analyze their speech production and how to become more actively involved in their own improvement. We have found that one of the most effective ways to develop this learner autonomy, especially in pronunciation, is with feedback that is visually demonstrable at the point of occurrence. In our experience, visual feedback has been crucial for helping learners both notice and analyze their errors. In this teaching tip, we explain how we use Praat as a visual-feedback tool for guided feedback practice in the classroom and for self-reflection homework.

INTRODUCTION

As pronunciation instructors, our goal is not to transform students into "accent-free" speakers but rather to help them improve control over their production of consonants, vowels, and prosodic elements to increase intelligibility – or, as we say, to strive for "approximation, not perfection" (Abercrombie, 1991 [1945]; Celce-Murcia, Brinton & Goodwin, 1996; Harmer, 1993; Jenkins, 1998; Kenworthy, 1987; Morley, 1991; Munro & Derwing, 1995; Neri, Cucchiarini & Strik, 2006; Seferoğlu, 2005). Many divergences from a native speaker (NS) standard—in grammar, word choice, fluency, and accentedness—do not cause communication difficulties; indeed, like many pronunciation instructors, we encourage our students to embrace these non-critical differences, both as a way of signaling to NSs their non-native speaker (NNS) status (which can provide social capital in the event of a speech mistake that could be considered serious or offensive if made by a NS) and as a form of multilingual and cultural pride. What actually constitutes an error depends to some degree on the individual student; learners from the same L1 background may share similar divergences, but the degree to which each student's intelligibility is negatively impacted will depend largely on the frequency (Munro & Derwing, 2006) and markedness (Santos, 1987) of their particular divergences, among other factors (Levis, 2007). Jenkins (2000, 2002) proposed a set of highly-salient features that require the greatest attention from both students and teachers, called the Lingua Franca Core (LFC), including categories such as high-frequency commonly conflated consonant sounds, coda consonant deletion, and voice quality, while sidelining less critical features such as /θ/ and /ð/, dark [ɫ], or vowel reduction, which, if misarticulated, do not typically cause significant comprehension difficulties for the listener. We generally follow Jenkins's recommendations, though with the caveat that, particularly given our students' ESL/EAP environment, and because many of them are training to become teachers, their self-confidence is also a significant factor for learning. So, if students wish to align themselves more closely with the standard for certain features that they feel, rightly or wrongly, are impacting their intelligibility, we find it worthwhile to first discuss with them the relative importance of those features for their particular context. Then, if they still feel strongly about it, we help them finesse their control of those features.

In this effort, we continually look for ways to effectively incorporate software, multimedia websites, and other technology into instruction. In our courses, students aim to approximate academic US English, so the resources we offer in this paper are tailored to that audience; however, if your students are targeting a different variety of English or if you teach a language other than English, you will of course want to use models or resources in that dialect or language.

Error correction relies on learner awareness, and it has long been agreed that the nature and timing of feedback can increase this awareness (Annett, 1969; Imber, Maynard & Meechan, 2012; Kulik & Kulik, 1988; Panova & Lyster, 2002). Ideally, then, we should provide feedback on students' pronunciation strengths and weaknesses when they occur, but without interrupting the flow of speech. In the absence of an oral equivalent of the "track changes" feature in Microsoft *Word*, this remains a challenging prospect.

Fortunately, the increasing ubiquity and variety of software designed for language learners is beginning to level the playing field. In our opinion, computer-assisted visual feedback is one of the most promising avenues for providing point-of-occurrence comments to help students improve intelligibility in their target dialect.

In pronunciation research specifically, visual feedback of various types has long been touted as a method of teaching intonation (Anderson, 1960; Cranen, Weltens, de Bot & van Rossum, 1984; de Bot, 1980, 1983; Hardison, 2004; Hermes, 1998; Hirata, 2004; James, 1979; Levis & Pickering, 2004; Ruellot, 2007; Seferoğlu, 2005; Taniguchi & Abberton, 1999; van Wieringen, 1994; Weltens & de Bot, 1984), perhaps because it cannot be seen by the naked eye in the same way that the articulation of segmentals is often observable (e.g., the position of the tongue, lip rounding). However, a non-expert can interpret a pitch contour representing intonation more intuitively than a spectrogram, making visual feedback a more natural fit for teaching intonation.

Other researchers – notably Brett (2004), Gonet (2001), Lambacher (1999), Landahl & Ziolkowski (1995), and Molholt (1990), – have successfully used computer-assisted visual feedback to teach segmentals, as well. Few, though, had advocated for using visual feedback to teach global pronunciation skills until Derwing, Munro & Wiebe (1998), who argued that students could be trained to improve their speaking rate, intonation, rhythm, vocal projection (volume and clarity), word stress, and sentence stress through visual feedback. Other studies have confirmed that visual feedback has led to improved perception and production of both segmentals and suprasegmentals (Hirata, 2004; Seferoğlu, 2005), particularly when students were able to visually compare their own output to a model speech sample (Hermes, 1998; James, 1979).

For visual feedback to be effective, it must include instructor monitoring, so that the student receives concrete feedback to help him or her interpret the acoustic display. As de Bot and Mailfert (1982) observe, "it became clear that a student could not just be placed in front of a display and told to learn correct intonation" (p. 72). Moreover, a number of researchers (Chun, 1989; Hirata, 2004; Neri, Cucchiarini, Strik & Boves, 2002; Pennington, 1999) agree that simply providing an acoustic display is unhelpful; the huge amount of raw acoustical data is overwhelming, and it provides no guidance to students on how to determine which features warrant focus, how to modify their speech to approximate a model, or how closely aligned with the model their speech must be in order to be considered “acceptable.” Watson and Kewley-Port (1989) agree: “The critical information necessary to modify erroneous speech productions should be presented as a salient feature of the display, rather than as one that must be sought in the midst of a complex array of irrelevant detail” (p. 37). Thus, instructors must help their students recognize which features of their acoustic display are relevant to their errors, and which differences do not affect listener understanding, so that they can eventually learn to study a display and pinpoint their personal key problem areas, specifically those that impact their intelligibility.

Of the various visual-feedback programs available, our favorite, *Praat* (Boersma & Weenink, 2017), was not originally intended for pedagogical purposes. In a nutshell, *Praat* is a robust speech analysis program developed by phoneticians for phoneticians who study features such as vowel duration, voice onset time, and formant frequencies. However, it can also be highly accessible to pronunciation instructors—even those with little experience—and, ultimately, to students. Despite the program’s complexity (and the potentially intimidating level of detail in its text-heavy interface), what drew us to *Praat* is that it also provides an extremely useful visual depiction of a student’s speech. Readers who have used *Audacity*[1] (Mazzoni & Dannenberg, 1999) may note similarities, although in our opinion, while *Audacity* is initially more intuitive and user-friendly, we have found *Praat* to be more useful for our purposes because it provides a number of features simultaneously. The standard *Praat* display (*Figure 1*) shows:

- a **waveform**, which illustrates volume (i.e., intensity) and the presence or absence of linking (i.e., connecting words without producing breaks between them) and pausing;
- a **spectrogram**, for pinpointing segmentals and word boundaries[2];
- a **pitch contour**, for measuring pitch range and identifying stressed syllables, which together contribute to intonation;
- and **timing bars**, for evaluating speaking time and vowel duration.

[1] This is another free speech recording and analysis software, which is available at http://www.audacityteam.org/.
[2] Though linking in speech makes it difficult (if not impossible) for a listener to distinguish discrete word boundaries, *Praat*’s visual display allows a user to at least roughly identify these delineations – though there can be significant overlap in the case of coarticulated phonemes – which can be useful, for example, in isolating a specific word or phrase that a student wishes to practice.

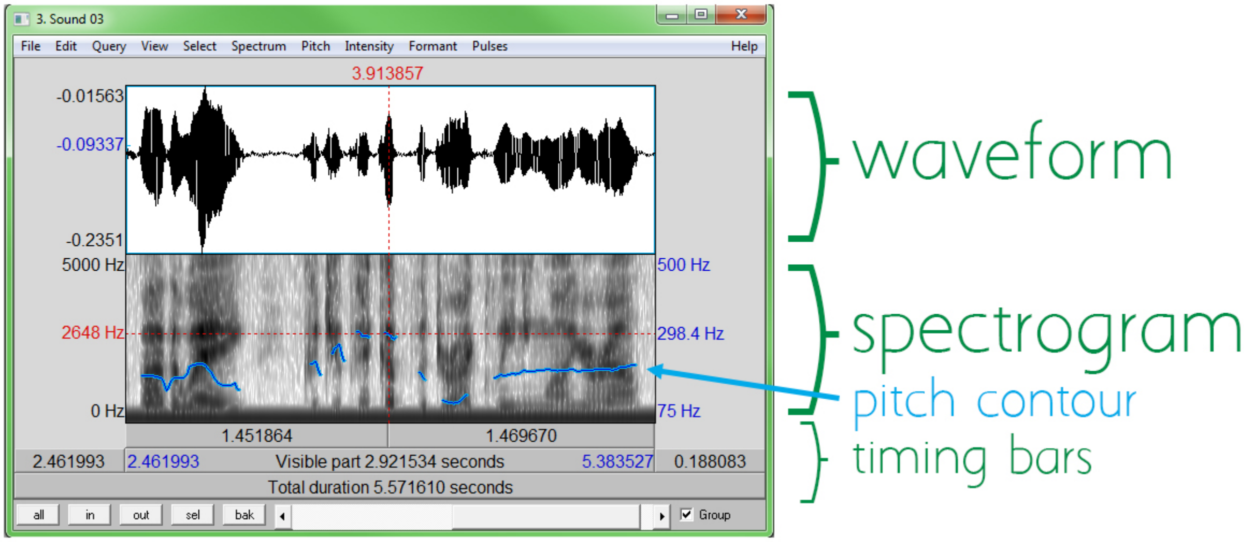


Figure 1. Praat display.

Students can analyze similarities and differences within these features when they compare their recording with the same utterance recorded by a native speaker. As we mentioned earlier, for non-experts (i.e., our students), spectrograms are less intuitive to interpret than the waveform, pitch contour and timing bars, but with guidance and practice, all four features can become useful resources for student improvement. A few examples are provided in figures 2-5 below.

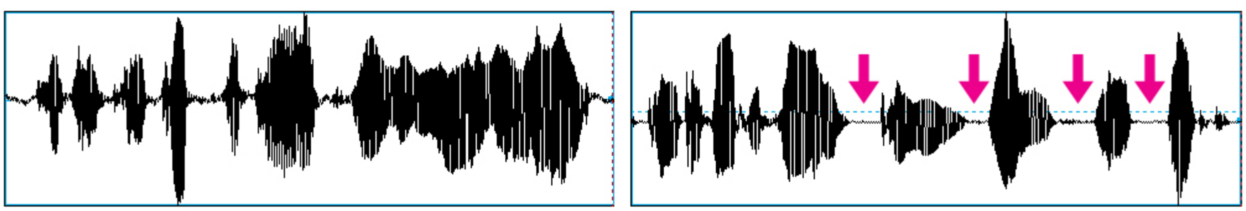


Figure 2. Linking comparison

Figure 2 shows part of the *Praat* waveform from Figure 1, for the phrase "Could you pick up the dry-cleaning while you're out?" The model, on the left, illustrates a well-linked phrase: even when the volume decreases between syllables, the waveform never completely flattens out (though it comes close during the unreleased [p˺] at the end of "up"). In contrast, the student recording, on the right, has four distinct un-linked breaks after the words *dry*, *cleaning*, *while* and *you're*, indicated with arrows.

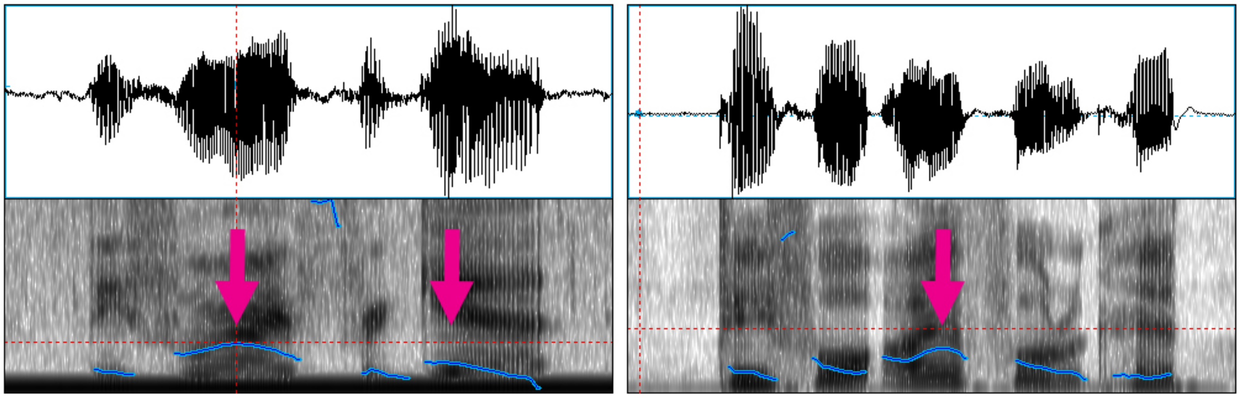


Figure 3. Stress comparison

Figure 3 shows a *Praat* display for the phrase “Before I forget…” The model recording, on the left, shows primary and secondary stress on the syllables *Before I forget*, indicated with arrows; the student recording, on the right, shows only primary stress on a different syllable: *Before I forget*. We typically ask students first to count how many stressed syllables they find in their recording vs the model, and then listen to each one to determine whether they’ve stressed the same syllables as the model. (This is most straightforward for formulaic expressions like the one above, where stress is relatively fixed. For other utterances, where a shift in stress might make no difference, create a difference in connotation, or sound unnatural, we discuss the appropriacy on a case-by-case basis.)

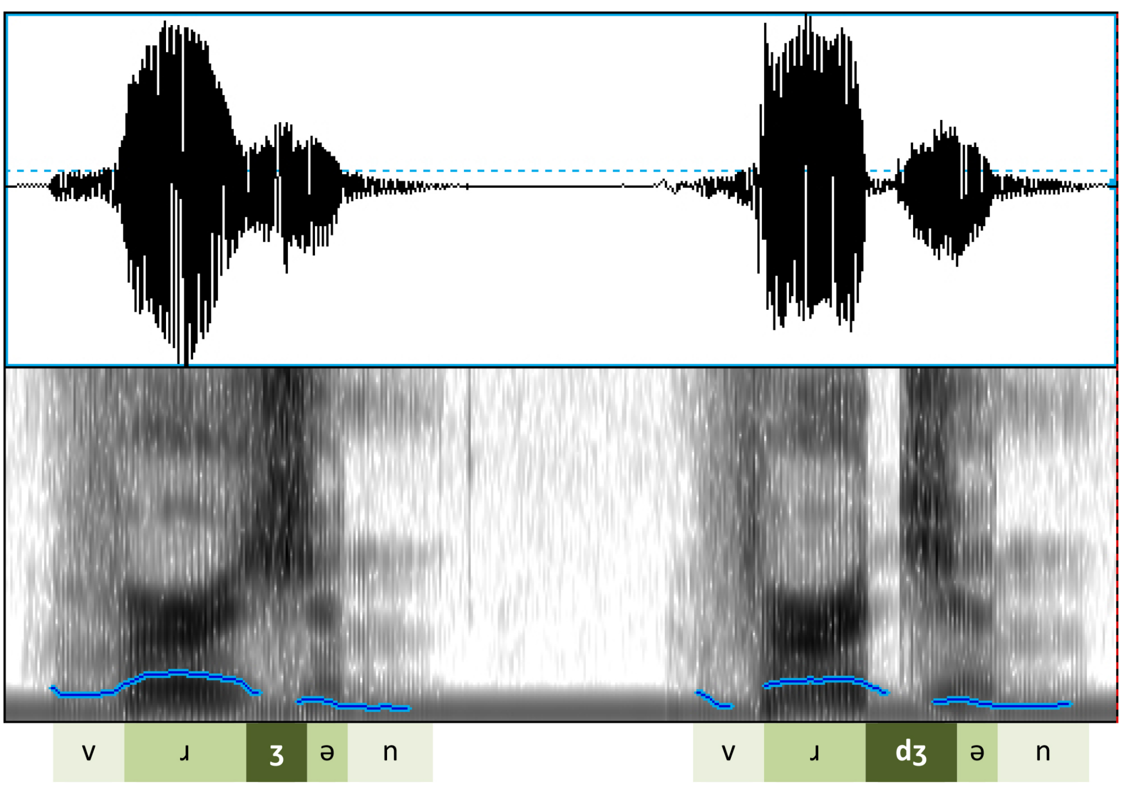


Figure 4. Consonant comparison

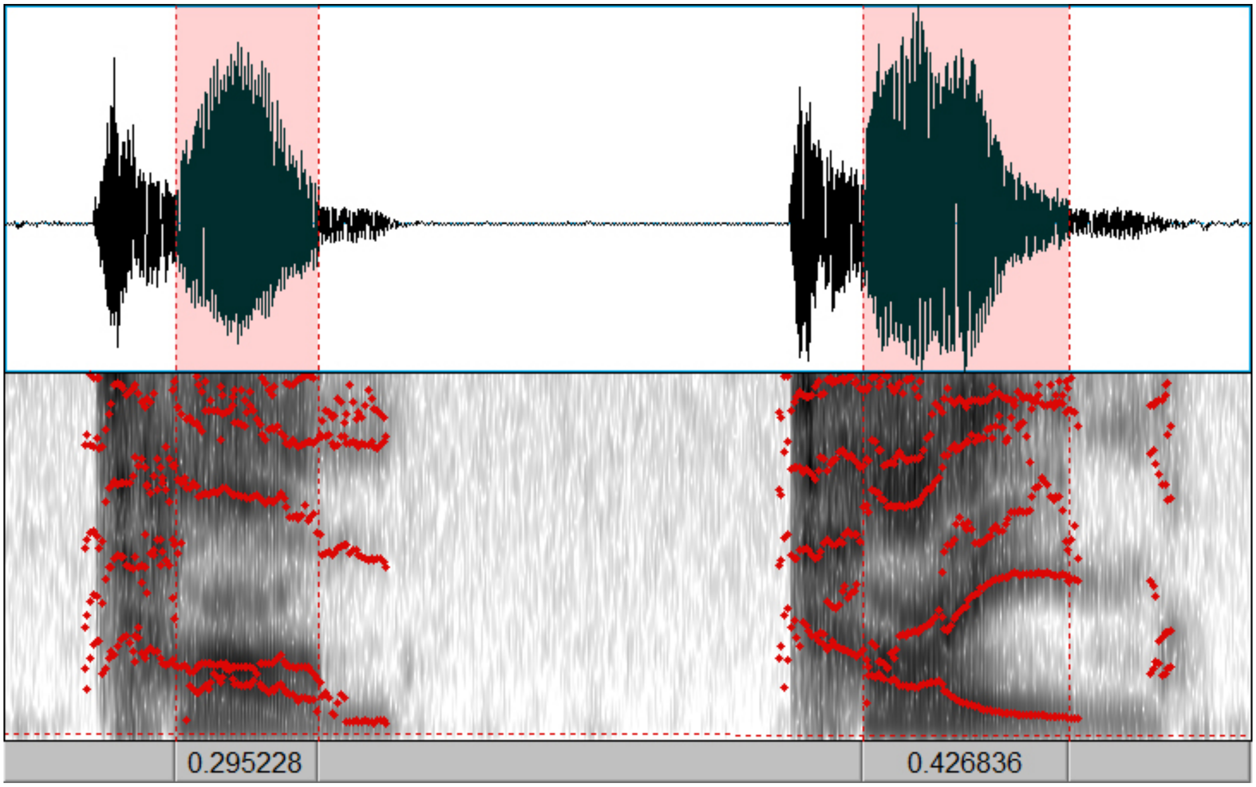


Figure 5. Vowel comparison

Figures 4 and 5 illustrate instances in which the *Praat* spectrogram can be helpful. Because teaching students to interpret spectrograms is labor-intensive and may not reap commensurate benefits, we prefer to use the spectrogram as a resource only when it clearly illustrates a sound difference that is causing significant problems. For example, in Figure 4, by showing the break caused by a stop consonant, *Praat* makes it evident whether a student is pronouncing fricative /ʒ/ (in *version*, left) or affricate /dʒ/ (in *virgin*, right). Similarly, our students often struggle not to monophthongize diphthongs, especially before nasals, which can create comprehension difficulties for pairs like *pain* and *pen*, or *ice* and… ahem. Figure 5 (a spliced image, with formant markers overlaid) illustrates how *Praat* can help students recognize whether they're pronouncing a shorter monophthong (as in *Tom*, left) or a longer diphthong (as in *time*, right, which shows a clearly shifting vowel). The spectrogram can also help to illustrate somewhat broader issues, like dropped final consonant clusters, deleted syllables, or epenthetic vowels. For more fine-grained distinctions, such as a conflation of /i/ and /ɪ/ in *gene* versus *gin*, /θ/ and /s/ in *math* versus *mass*, or /r/ and /l/ in *right* versus *light*, the spectrographic analysis becomes more cumbersome than we feel is worth our students' time and effort.

As always, we remind students that they need not match the model exactly, merely work to approximate the specific features that might cause confusion (in our experience, the most common features needing attention are linking, dropped syllables, epenthetic vowels, pitch, and stress placement). For example, baseline pitch is highly individualistic, correlating largely to vocal tract length, which can vary significantly by gender (Childers & Wu, 1991; Rendall, Kollias, Ney & Lloyd, 2005). We find that our students tend to produce relatively flat intonation, so, rather than trying to match the model speaker's exact pitch, we encourage students to match the pitch *range* (as discussed in Appendix 1), which helps them more closely approximate native-speaker intonation.

Before computer-assisted visual feedback became readily available, self-studying students more commonly listened to audiocassettes or CDs in a language lab, having to rely on their intuition to guess whether they were pronouncing a word or phrase correctly. The opportunity to work with audio-visual depictions of their errors or strengths without interrupting the natural flow of speech creates a hands-on experience for the learner, and being able to see their speech displayed on the monitor gives students an immediate and quantifiable way to assess their pronunciation. Best of all, *Praat* is free to download, so students can use it outside the classroom at no expense.

It should be mentioned that, compared to the common visual shorthand of Microsoft Office, for example, *Praat* has a sharper learning curve for both instructors and students. As can be seen from Figure 1, the visual displays are packed full of information, so students require training in order to get comfortable using the program and even more training to learn to analyze their speech appropriately.

DISCUSSION

First, we provide an hour-long orientation at the start of the course, to introduce students to the specific functions and features that will be relevant and helpful. In order to provide further scaffolding, we encourage frequent usage of *Praat* both in and out of class, so that students become comfortable using it.

Our ultimate goal is for students to be able to use *Praat* for autonomous self-study, so in the weeks following the orientation, we help prepare students to "fly solo" by getting them gradually accustomed to using the program. During a ten-week course, we typically meet students once or twice a week for an hour. Throughout these confidence-building weeks, the three most helpful factors are instructor support, guided practice time, and an accountability procedure.

First, **instructor support** is essential because as noted, the displays are information-dense. We need to train students in what to look for, how to interpret what they're seeing, and then how to reflect on what modifications they might make.

Second, during **guided practice time**, the instructors facilitate the process of analyzing and providing feedback. To start, however, we must create a "sound bank" of model sentences with which the students can practice. We ask each student to invent their own sentence, incorporating formulaic expressions (from the Verbal Stratagems website in Table 1 below), terms from their field or from daily life, and words they have trouble pronouncing. We record a model version of each sentence, and the student then records their own version before downloading the model and comparing the two, as described above. This activity not only enables the whole class to get involved, but also generates a set of model sentences which students can use for later practice at home.

We continue with *instructor-led* analysis, in which we sit down with each student to help them compare and analyze their recordings. (For this reason, we like to hold our classes in a computer lab, so that we can spend at least the last 10 to 15 minutes working with students.)

In time, we transition to *pair or group-led* analysis, in which students analyze and discuss each other's recordings before reporting out to the class. To analyze their recordings, we can either move individually from student to student, or ask a particularly brave student to share their recording for the whole class to view and discuss.

Once students are comfortable and successful at analyzing without our help, we move to *independent* analysis, encouraging them to practice at home on their own. Because instructor modeling is critical for student improvement, we felt it was vital to make sure that students understood the software, had a chance to practice it with the instructors, and were comfortable using it before we asked them to practice at home and self-monitor.

Finally, these independent homework assignments are the core of our **accountability procedure**. We assign a set of formulaic expressions that students must expand into full sentences, as in our guided practice time activity, which they then record and submit. We urge students to *micro-practice,* i.e., practice of about 2-3 minutes, 5 times per day, because this spaced approach is far more effective for improving recall and building muscle memory than longer but less frequent sessions (deWinstanley & Bjork, 2002; Maley, 2017; Rohrer & Pashler, 2010; Wiklund-Hörnqvist, Jonsson & Nyberg, 2013). Additionally, breaking up the time into shorter chunks is not only more manageable in a busy schedule, but it also helps students to regard their practice time as motivational and fun rather than exhausting and boring.

We frequently found that students were engaged and enthusiastic about *Praat* in class, but still felt unsure about how exactly to use it without an instructor present, so we created a "Raising Awareness" activity (Appendix 1). Its purpose is two-fold: initially to train our students to use *Praat*, and ultimately as a step-by-step guide they can use for working independently.

In order to practice at home, students will need model sound files of their target dialect for comparison. Ideally, we recommend that you record the model files for your students yourself. If you do not have time, you lack a good recording environment, you want students to hear a different voice, you are not a speaker of your students' target dialect, or if your students want further practice materials after your course has ended, you could point your students to the short sample list of reliable .wav archives (of primarily US and UK English) provided in Table 1, where your students can download sound files for *Praat* that have been captured, for the most part, from TV and film. However, do alert students to avoid sound files that contain background noise, which muddies the acoustic display.

Table 1

WAV archives

Selected Internet Resources	
Praat	***http://www.praat.org/***
The Daily .WAV	*http://www.dailywav.com/*
The Frogstar Huge WAV Archive	*http://www.frogstar.com/wav/*
MovieWAVs	*http://www.moviewavs.com/*
WAVCentral	*http://wavcentral.com/*
WAVSource	*http://www.wavsource.com/*
UM ELI Verbal Stratagems (no WAV files)	*http://www.umich.edu/~flcourse/verbstrat/*

Figures 6, 7, and 8 below illustrate the potential for student improvement using *Praat*. The student whose acoustic displays are presented, LS, enrolled in a seven-week *Praat* workshop (as described in Imber, Maynard, Ohlrogge & Chien, 2009) for one hour a week. The workshop consisted of brief lessons on the topics from Appendix 1 followed by time for the students to practice with *Praat* in class. They were also required to keep a homework log to track how frequently they practiced on their own. Figure 6 shows the *Praat* display for one of the model sentences they were given: “By the way, could you pick up the dry cleaning while you’re out?”

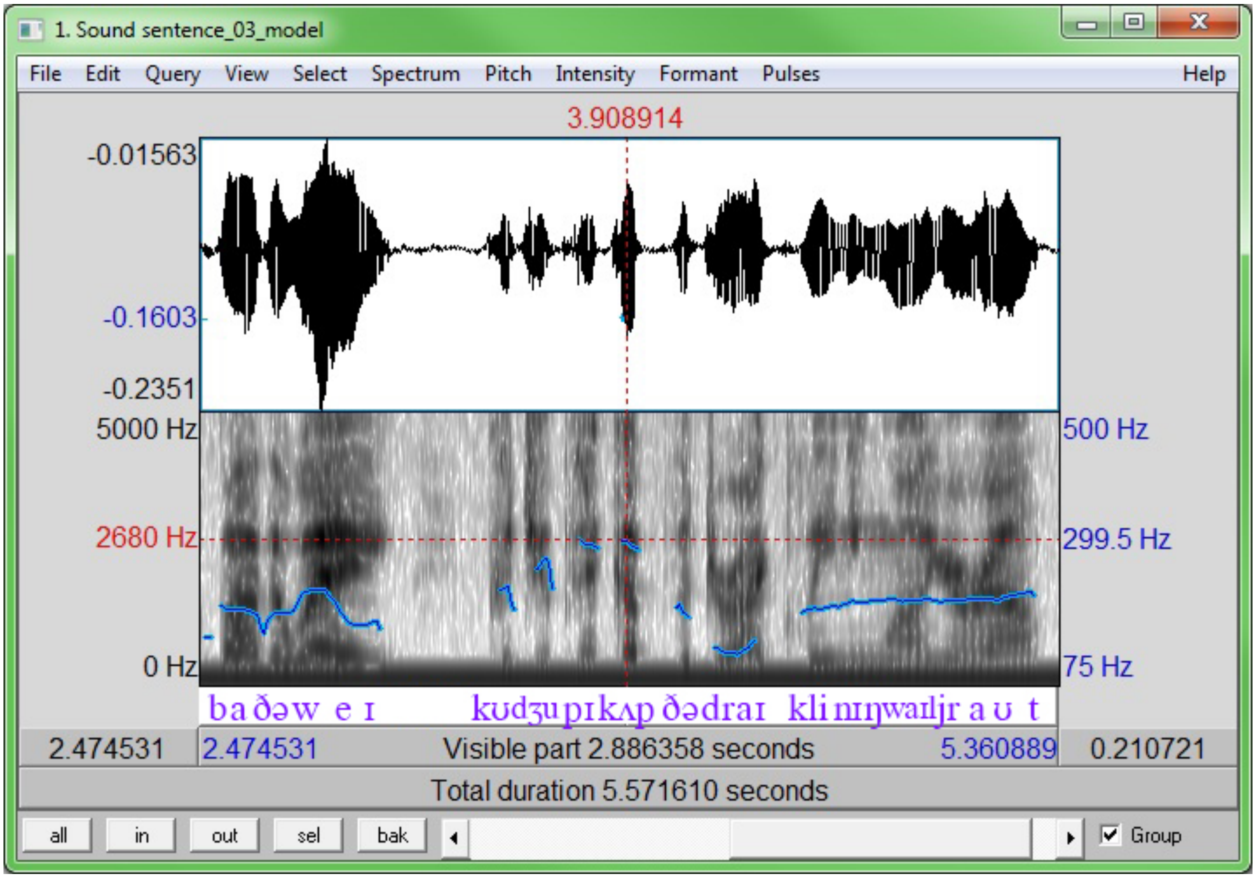


Figure 6. Model sentence

It is important to note that, as seen in Figure 6, the pitch contour does not display for voiceless sounds (e.g., Anderson-Hsieh, 1993), so the pitch contour may show breaks even if no pause is present in the recording. If learners have a particular need for pitch or intonation improvement, you may prefer to use model sentences with mainly voiced sounds. Another important point to note is that, as seen mildly in Figure 6 and dramatically in Figure 8, extraneous noises in the audio can produce a misleading swerve in the pitch contour, sometimes even causing it to jump to the extreme upper or lower limit of the spectrogram. While this is not a frequent issue, or a major drawback since the veracity of any apparent pitch changes can be easily double-checked by listening to the audio, it can be minimized by making your recordings in a location as soundproof and free of background noise as possible. Our location, evidently, was not: these sentences were recorded in a university computer lab.

As described in Appendix 1, students practicing with a model sentence are asked to compare their own speed, pausing and linking, pitch, stress, intonation, and pronunciation of segmentals to the model, making note of any significant similarities or differences. While students are welcome to work on any or all of the features that catch their attention, we encourage them to focus on, at most, their top 3 or 4 problem areas (assuming they have that many). Not only does this make their task more manageable, it also helps them concentrate their efforts on features that affect their intelligibility, rather than spending their time on "accent reduction" of non-critical differences.

LS recorded this sentence, as part of a 6-sentence diagnostic set, twice: during the first-day orientation session (Figure 7) and again on the last day of the workshop (Figure 8). During the intervening six weeks, LS practiced with an additional 30 sentences, but was not exposed to the six diagnostic sentences again during that time period, nor did she ever view or listen to the model sound files for those sentences.

Each of the six diagnostic sentences, and the thirty subsequent practice sentences, were composed of two parts: a formulaic expression, followed by a colloquial English sentence. With the colloquial English sentence, we hoped to avoid giving students a sentence they might have seen before. Our hope was that, after using visual feedback to practice with other colloquial English sentences for six weeks, the students would have gotten a feel for the "flavor" of features like English-language linking and intonation. We used formulaic expressions to introduce these sentences because formulaic speech is much more "fixed" in terms of pausing, linking and intonation, as well as grammar; therefore, these phrases provided a less ambiguous target than non-formulaic utterances, and gave students an advantage in improving overall intelligibility.

The needs of any given student will of course be individualistic, largely influenced by their L1 and the degree to which any divergence from the model interferes with their intelligibility. A major benefit of *Praat* is that it facilitates this individualistic approach to error feedback. In the case of LS (Figure 7), her phonology, though noticeably accented—for example, [zə vei] for *the way*, [klinɪŋk] for *cleaning*—did not pose problematic barriers to her intelligibility. More salient were the amount of time it took her to produce the utterance and her intonation. LS's speed of 5.51 seconds (shown at **1** in Figure 7) was nearly twice as long as the model, at 2.89 seconds. The source of this discrepancy was largely to be found in her intonation (**2**), which was characterized by consistent rise-fall stress on almost every word ("**Could** you **pick** up the **dry clea**ning **while you're out**?"), even on the final word of the question (**3**), with frequent brief pauses.

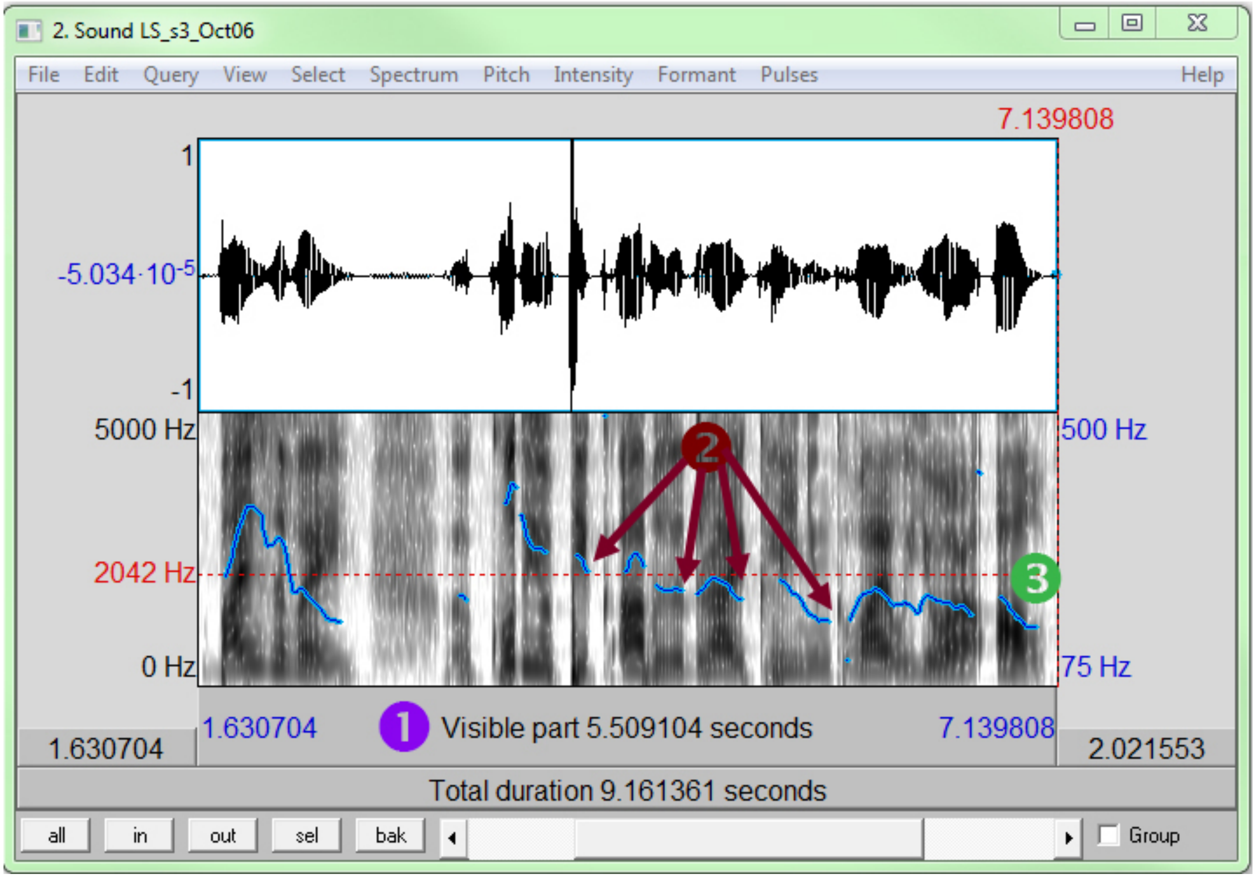


Figure 7. LS initial diagnostic

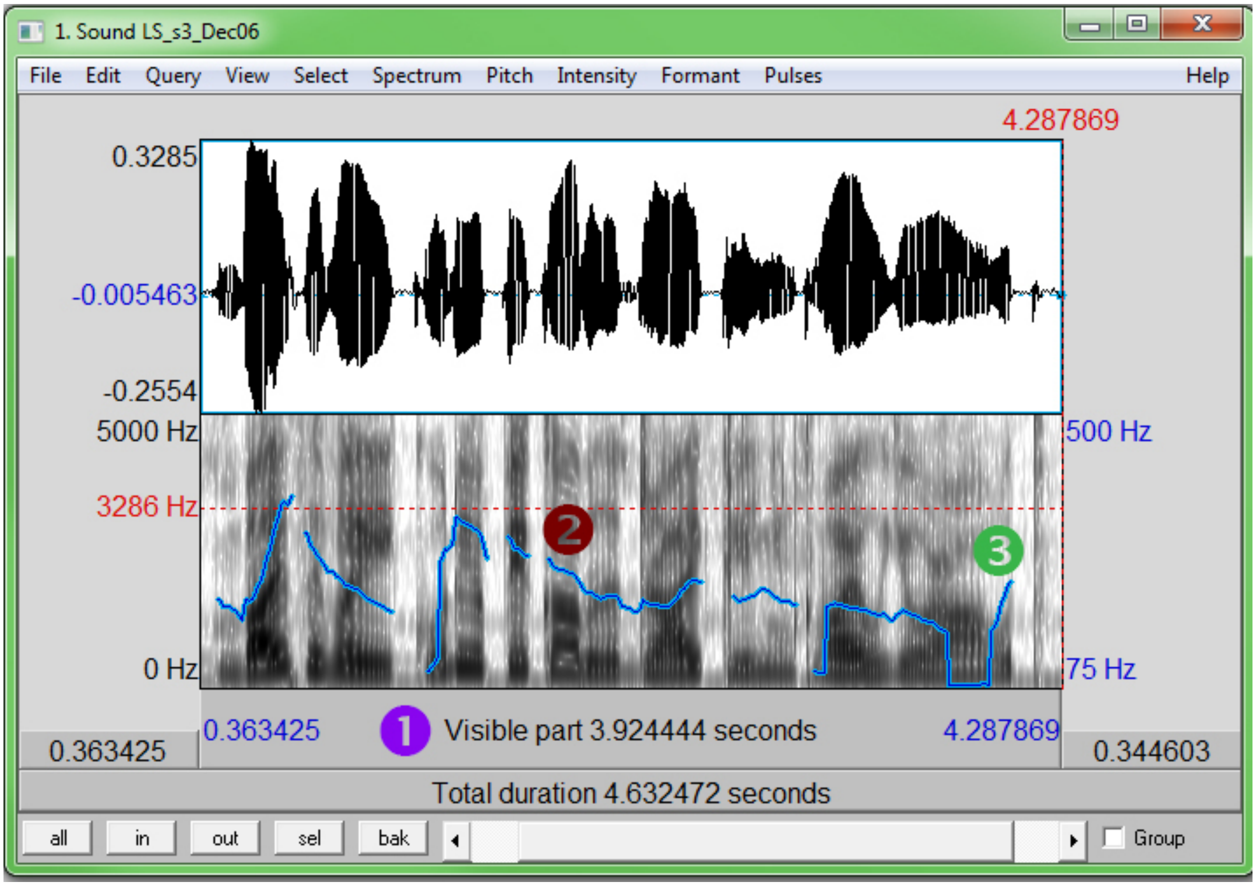


Figure 8. LS final diagnostic

From the visual display alone, we can see that after six weeks of practice, **(1)** LS's pace more closely approximates the model sentence (now 3.92 seconds, down from 5.51), and her intonation pattern has improved, **(2)** becoming smoother (stressing only the key words, and eliminating the frequent pauses), and **(3)** with phrase-final rising (rather than falling) intonation appropriate for yes/no questions. We would expect this more native-like intonation, in particular, to improve LS's intelligibility, given that research has demonstrated that unfamiliar (or unvarying) intonation contours can slow down or decrease comprehension (Braun, Dainora & Ernestus, 2011; Hillenbrand, 2003; Holub, 2010).

LS's progress was notably dramatic, even among her workshop peers. As with most pronunciation instruction, students typically see results with *Praat* after a longer period of regular practice. In general, compared with the control group, we saw mild improvement after six weeks of practice with *Praat*. Of course, for more pronounced improvement, most students will need a longer timespan, and we cannot overstate the importance of frequency of use.

Students should also note that it is difficult to improve segmentals with *Praat* in a self-study context, unless they can read spectrograms; otherwise, without visual support, they are entirely reliant on perception. Even in the classroom, students require a fair amount of instructor support for comparison and analysis of segmentals. The waveforms and pitch contour, however, are dramatically informative about the suprasegmentals – linking, stress, pitch and intonation – because these elements are relatively easy to interpret in waveforms and spectrograms, making it much easier for students to compare and analyze prosody. But it is important to reiterate that not every marker of foreign accent or language control affects intelligibility (Derwing & Munro, 1997); only those divergences that interfere with comprehension need to be targeted for improvement. Although student goals tend to align with their intelligibility needs, if the instructor feels that a student has chosen a pronunciation goal that does not interfere with listener comprehension, this goal can be discussed with the student.

Finally, language level plays a role in how (or whether) students benefit from *Praat*. It is important that students' language skills be strong enough both to understand the instructions and to navigate the program, but not so advanced that they have little room for significant improvement. For example, at the University of Michigan, we have seen more pronounced results with visiting international scholars, who are working or studying in the USA temporarily, than with international graduate students, who had to meet a minimum TOEFL score for admission to a four-year degree program. Still, even this more advanced group often has considerable room for improvement, especially those who are aiming to become teaching assistants.

CONCLUSION

Both empirical research, which showed a moderate effect on fluency and intonation after seven weeks (Imber et al., 2009), and anecdotal comments from our students suggest that the use of visual feedback enhances their language-learning experience, not only in terms of measurable improvement but also self-confidence. As one of our students remarked, "If I can see it, I can say it, I can practice." Students appreciate being able to see differences, such as the "melody" of English, between their speech and the model, and value the ability to practice as much as they want without bothering the "speaker." While some training is necessary, both to accustom students to using the program and to focus only on the errors that are significantly impacting their intelligibility, we have found it to be time well spent, because quality visual feedback empowers students to self-assess their speaking skills with more certainty, less guesswork, and greater impact on their comprehensibility. Overall, the incorporation of a visual feedback tool such as *Praat* can have a positive impact on the language learning classroom for students interested in improving pronunciation and increasing intelligibility.

ABOUT THE AUTHORS

Brenda Imber, PhD, has taught at the University of Michigan English Language Institute (ELI) for over 25 years. Carson Maynard, MA, has worked at the ELI since 2003. Their professional interests include pronunciation, teacher training, and integrating technology into language learning.

Maria Parker, MS, has taught at Duke University English for International Students for over 25 years, and served as program director from 2000 – 2015. Parker and Imber co-authored *Integrated Lessons: Pronunciation & Grammar* (University of Michigan Press, 1993).

REFERENCES

Abercrombie, D. (1991 [1949]). Teaching pronunciation. In A. Brown (Ed.), *Teaching English Pronunciation: A Book of Readings* (87 – 95). London: Routledge.

Anderson, F. (1960). An experimental pitch indicator for training deaf scholars. *Journal of the Acoustical Society of America*, 32 (18), 1065 – 1074.

Anderson-Hsieh, J. (1993). Interpreting visual feedback on suprasegmentals in computer assisted pronunciation instruction. *CALICO Journal*, 11 (4), 5 – 22.

Annett, J. (1969). *Feedback and human behaviour: The effects of knowledge of results, incentives and reinforcements on learning and performance.* Oxford, England: Penguin Books.

Boersma, P., & Weenink, D. (2017). Praat: doing phonetics by computer (Version 6.0.25) [computer program]. Available from http://www.praat.org/

Braun, B., Dainora, A., & Ernestus, M. (2011). An unfamiliar intonation contour slows down online speech comprehension. *Language and Cognitive Processes*, 26 (3), 350 – 375.

Brett, D. (2004). Computer generated feedback on vowel production by learners of English as a second language. *ReCALL: Journal of Eurocall*, 16 (1), 103 – 113.

Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (1996). *Teaching Pronunciation*. Cambridge, England: Cambridge University Press.

Childers, D. G., & Wu, K. (1991). Gender recognition from speech. Part II: Fine analysis. *The Journal of the Acoustical Society of America*, 90, 1841 – 1856.

Chun, D. M. (1989). Teaching tone and intonation with microcomputers. *Computer Assisted Language Instruction Consortium (CALICO) Journal*, 7 (1), 21 – 46.

Cranen, B., Weltens, B., de Bot, K., & van Rossum, N. (1984). An aid in language teaching: the visualization of pitch. *System*, 12 (1), 25 – 29.

de Bot, K. (1980). The role of feedback and feedforward in the teaching of pronunciation – an overview. *System*, 8, 35 – 47.

de Bot, K., & Mailfert, K. (1982). The Teaching of Intonation: Fundamental Research and Classroom Applications. *TESOL Quarterly*, 16 (1), 71 – 77.

de Bot, K. (1983). Visual Feedback of Intonation I: Effectiveness and Induced Practice Behavior. *Language and Speech*, 26 (4), 331 – 350.

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., and Wiebe, G. (1998). Evidence in Favor of a Broad Framework for Pronunciation Instruction. *Language Learning*, 48 (3), 393 – 410.

deWinstanley, P. A., & Bjork, R. A. (2002). Successful lecturing: Presenting. information in ways that engage effective processing. In Halpern, D. F., & Hakel, M. D. (Eds.), *Applying the Science of Learning to University Teaching and Beyond* (19 – 31). San Francisco, CA: Jossey-Bass.

Elliott, H. (2016). The Daily .WAV. Retrieved from http://www.dailywav.com/

Goebel, J. (2016). Frogstar.com: The Audio Collections. Retrieved from http://www.frogstar.com/wav/

Gonet, W. (2001). Obstruent Voicing in English and Polish: A Pedagogical Perspective. *International Journal of English Studies*, 1 (1), 73 – 92.

Hardison, D. M. (2004). Generalization of Computer-Assisted Prosody Training: Quantitative and Qualitative Findings. *Language Learning & Technology*, 8 (1), 34 – 52.

Harmer, J. (1993). *The practice of English language teaching*. New York, NY: Longman.

Hermes, D. J. (1998). Auditory and visual similarity of pitch contours. *Journal of Speech, Language and Hearing Research*, 41 (1), 63 – 72.

Hillenbrand, J. M. (2003). Some effects of intonation contour on sentence intelligibility. *The Journal of the Acoustical Society of America*, 114 (4), 2338 – 2338.

Hirata, Y. (2004). Computer assisted pronunciation training for native English speakers learning Japanese pitch and durational contrasts. *Computer Assisted Language Learning*, 17 (3 – 4), 357 – 376.

Holub, E. (2010). Does intonation matter? The impact of monotony on listener comprehension. *The Interpreters' Newsletter*, 15, 117 – 126.

Imber, B. P., Maynard, C., Ohlrogge, A., & Chien, Y-H. (2009). Using visual feedback to improve L2 pronunciation. In Katsampoxaki-Hodgetts, K. (Ed.), *Options & Practices of LSP Practitioners Conference Proceedings* (408 – 441). Heraklion, Crete: University of Crete Publications.

Imber, B. P., Maynard, C., & Meechan, P. (2012). An approach to video annotation for oral skills feedback. In Pattison, T. (Ed.), *IATEFL 2011 Brighton Conference Selections* (212 – 215). Eynsham, UK: Information Press.

Imber, B. P., Maynard, C., Parker, M., Reilly, M., Ohlrogge, A., Chien, Y-H.. Meechan, P., Cameron, P., & Beals, J. (2017). Verbal Stratagems. Retrieved from http://www.umich.edu/~flcourse/verbstrat/

James, E. F. (1979). Intonation through visualization. In Hollien, H. & Hollien, P. (Eds.), *Current Issues in the Phonetic Sciences* Vol. IV, Amsterdam Studies in the Theory and History of Linguistic Science (295 – 301). Amsterdam: John Benjamins.

Jenkins, J. (1998). Which pronunciation norms and models for English as an international language? *ELT Journal*, 52, 119 – 126.

Jenkins, J. (2000). *The Phonology of English as an International Language: New Models, New Norms, New Goals.* Hong Kong: Oxford University Press.

Jenkins, J. (2002). A sociolinguistically-based, empirically-researched pronunciation syllabus for teaching English as an international language. Applied Linguistics, 23, 83-103.

Kenworthy, J. (1987). *Teaching English pronunciation*. New York, NY: Longman.

Kulik, J. & Kulik, C. (1988). Timing of feedback and verbal learning. *Review of Educational Research*, 58 (1), 79 – 97.

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.

Landahl, K. L. & Ziolkowski, M. S. (1995). Discovering phonetic units: Is a picture worth a thousand words? *Papers from the Regional Meetings, Chicago Linguistic Society*, 31 (1), 294 – 316.

Levis, J. (2007, March). *Guidelines for promoting intelligibility.* Paper presented at the meeting of International TESOL, Seattle, WA.

Levis, J., & Pickering, L. (2004). Teaching intonation in discourse using speech visualization technology. *System*, 32, 505 – 524.

Maley, A. (2017, April). *Is what we know, what we do?* Paper presented at the meeting of International Association of Teachers of English as a Foreign Language (IATEFL), Glasgow, UK.

Mazzoni, D., & Dannenberg, R. (1999). Audacity (Version 2.1.3) [computer program]. Available from http://www.audacityteam.org/

Molholt, G. (1990). Spectrographic analysis and patterns in pronunciation. *Computers and the Humanities*, 24, 81 – 92.

Morley, J. (1991). The pronunciation component in teaching English to speakers of other languages. *TESOL Quarterly*, 25, 481 – 519.

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.

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.

Neri, A., Cucchiarini, C., & Strik, H. (2006). Selecting segmental errors in non-native Dutch for optimal pronunciation training. *International Review of Applied Linguistics in Language Teaching (IRAL)*, 44 (4), 357 – 404.

Panova, I., & Lyster, R. (2002). Patterns of corrective feedback and uptake in an adult ESL classroom. *TESOL Quarterly*, 36 (4), 573–595.

Pennington, M. C. (1999). Computer-aided pronunciation pedagogy: Promise, limitations, directions. *Computer Assisted Language Learning*, 12 (5), 427 – 440.

Rendall, D., Kollias, S., Ney, C., & Lloyd, P. (2005). Pitch (F0) and formant profiles of human vowels and vowel-like baboon grunts: The role of vocalizer body size and voice-acoustic allometry. *The Journal of the Acoustical Society of America*, 117, 944 – 955.

Rohrer, D., & Pashler, H. (2010). Recent research on human learning challenges conventional instructional strategies. *Educational Researcher*, 39, 406 – 412.

Ruellot, V. M. (2007). French pronunciation learning and computer-mediated visual feedback. *Dissertation Abstracts International, A: The Humanities and Social Sciences*, 67 (07), 2556.

Santos, T. (1987). Markedness theory and error evaluation: an experimental study. *Applied Linguistics*, 8 (3), 207 – 218.

Seferoğlu, G. (2005). Improving students' pronunciation through accent reduction software. *British Journal of Educational Technology*, 36 (2), 303 – 316.

Taniguchi, M., & Abberton, E. (1999). Effect of interactive visual feedback on the improvement of English intonation of Japanese EFL learners. *Speech, Hearing and Language*, 11, 76 – 89.

van Wieringen, M., & Abberton, E. (1994). The use of computerized representations in L2 acquisition of intonation: a pilot study. *Speech, Hearing and Language*, 8, 245 – 258.

Watson, C. S., & Kewley-Port, D. (1989). Advances in computer-based speech training: Aids for the profoundly hearing impaired. In McGarr, N. S. (Ed.), *The Volta Review*, 91, 29 – 45.

WavSource.com: Free Wav Files and Sound Bites. (2016). Retrieved from http://www.wavsource.com/

Weltens, B., & de Bot, K. (1984). The visualization of pitch contours: some aspects of its effectiveness in teaching foreign intonation. *Speech Communication*, 3, 157 – 163.

West, G. (2016). The MovieWavs Page. Retrieved from http://www.moviewavs.com/

White, B. (2016). Free Movie and TV Clips at Wav Central. Retrieved from http://wavcentral.com/

Wiklund-Hörnqvist, C., Jonsson, B., & Nyberg, L. (2013). Strengthening concept learning by repeated testing. *Scandinavian Journal of Psychology*, 55, 10 – 16.

Appendix 1: "Raising Awareness" activity

Guided Practice Exercise: Raising Awareness

1. **Load** Praat.
2. **Open** two files:
   i. Any model sound file
   ii. Your own recording of the same sound file
3. **Line them up** on your screen so that the model file is above your sound file.
4. Under *View > Show Analyses*, make sure that the **spectrogram** and **intonation line** are showing, and "formants" and "pulses" are **un-checked**. You can also increase the "Longest analysis" setting if the file is too long to display a spectrogram.
5. **Compare** your file to the model, and **look for** the following features:
   i. Speed
   - ☐ Is your sentence significantly longer or shorter than the model?
   - ☐ If so, can you find any places where your speed was much faster or much slower?
   - ☐ Can you tell what might be causing the difference? (If not, think about this as you analyze.)

   Look for: *vowel sound duration, dropped words*.

   ii. Pausing and Linking
   - ☐ Do you have the same number of pauses (breaks) as the model?
   - ☐ Are they roughly the same length?
   - ☐ Are they in the same place? i.e., are your thought groups (chunks) different from the model?
   - ☐ Do any of your pauses sever the links between words, or is your linking similar to the model?

   Look for: *micro-pauses, dropped final consonants (some, door, feel, etc), vowel insertion, voiced pauses ("big-uh dog"), fillers (uh, um, well, etc)*.

   iii. Stress, Pitch and Intonation
   - ☐ How many words did you stress, compared to the model?
   - ☐ Are there any phrases where your stress is on a different word?
   - ☐ Are there any words where your stress is on the wrong syllable?
   - ☐ How do your highest and lowest pitch compare to the model?
   - ☐ Do you have a wide pitch range, or does your pitch stay "flat"?
   - ☐ Do you see any phrases where your intonation was very different from the model?

   Remember that **the *blue line*** shows **pitch, stress and intonation**.
   If you click on **the peak or valley** of a syllable, **the *red number*** on the right (Hz) shows **how high or low** your pitch is.

   iv. Pronunciation
   - ☐ Are there any places where your words sound very different from, or similar to, the model?

   Look for: *problematic sounds (r/l, th, sh/ch/zh, etc)*.

Martin, I.A. (2017). Teaching pronunciation through homework assignments: The method of iCPRS. 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. 214-221). Ames, IA: Iowa State University.

TEACHING TIPS

TEACHING PRONUNCIATION THROUGH HOMEWORK ASSIGNMENTS: THE METHOD OF iCPRS

Ines A. Martin, Pennsylvania State University

While many previous classroom-based research studies aim at teaching pronunciation in a classroom environment, there is not always time in a lesson plan to include in-class pronunciation instruction. This teaching tip therefore focuses on teaching pronunciation through homework-based assignments called *innovative Cued Pronunciation Readings* (iCPRs; see also Tanner & Landon, 2009). This computer-delivered method of pronunciation instruction employs easily accessible technology, Microsoft *PowerPoint*, and has been shown to be effective in improving learners' L2 pronunciation skills in a pilot study with 22 first semester learners of German (Martin, 2015). Each iCPR unit is designed to require about 10 minutes of work. There are two kinds of iCPR units: perception and production training units. Both types of units contain native-speaker recordings of individual words. Perception training units then consist of two tasks: an accentedness detection task and a sound discrimination task. Production training units start with explicit instruction on the targeted sound and progress to active practice, for which the learners have to repeat individual words after a native speaker recording. At the end of a production unit, the learners have to record all words they practiced with recording software like *Audacity* and upload their productions to a course management system dropbox.

INTRODUCTION

While pronunciation has been shown to play an important role in effective communication (Celce-Murcia, Brinton, Goodwin, & Griner, 2010) and many previous studies have found that teaching pronunciation in a classroom setting helps to improve pronunciation proficiency among students (see Thomson & Derwing, 2015, for a narrative review), L2 pronunciation training is still often neglected in the foreign language classroom. Frequently stated reasons for this shortcoming are that many teachers do not feel comfortable with or adequately prepared to teach pronunciation and that in a full curriculum and system of standardized tests, there is no time to address pronunciation in the classroom (Breitkreutz, Derwing, & Rossiter, 2001; Derwing, 2013; Foote, Holtby, & Derwing, 2011; O'Brien, 2004). Furthermore, when directly asking teachers why they do not include pronunciation training in their classrooms, they often mention that singling out a student for a pronunciation mistake is perceived as harsher than, for example, correcting a student's syntax, which in turn makes teachers hesitant to single out students in front of the class. In this teaching tip, I therefore want to present a method of teaching pronunciation through homework assignments. Providing pronunciation training through homework assignments solves the problems mentioned above: it does not take up valuable in-class time, students are not being singled out in front of the class but rather get to practice in the comfort of their own home, and even those teachers that do not feel adequately prepared to teach pronunciation can assign the exercises as homework.

THE METHOD OF iCPRs

Background

The homework-based method of teaching pronunciation that I present in this teaching tip is called *innovative Cued Pronunciation Readings* (iCPRs). This method is computer-delivered, but uses easily accessible technology (i.e., Microsoft *PowerPoint*®) that every student can access from a home or library computer. The method is based on *Cued Pronunciation Readings*, which were first introduced by Tanner and Landon (2009). It differs from the original version in that Tanner and Landon addressed pronunciation only on the level of sentence melody and stress, whereas iCPRs more generally target any L2 sound or suprasegmental feature that has been shown to impede learners' intelligibility in classroom learning.

General Design

Each iCPR unit is designed to take about 10 minutes for students to complete. That means that the pronunciation homework does not have to be the only homework assigned on a given day, but that it can easily be combined with other, more traditional, homework assignments. Usually, I assign three iCPR units for one pronunciation focus, e.g. the acquisition of a new vowel or consonant. The first of these three units is always a perception unit (see below for a detailed description), followed by two units focusing on production skills. This order is based on previous research that has shown that perception may precede production in L2 pronunciation (see Thomson, 2011, for a review). This means that second language learners often first learn to correctly perceive an L2 speech sound before they can accurately produce it. Often, when hearing a nonnative speech sound, learners assimilate the sound's features to a sound that is part of their native phonetic inventory and therefore cannot hear the difference. By training them to perceive the difference, however, we lay the groundwork to allow them to produce the sound more accurately (Flege, 1995; Thomson, 2011).

Another key feature in the design of iCPR units is embedded native-speaker recordings that serve as models for learners to imitate. Ideally, these recordings should be completed by several native speakers to account for findings in high variability phonetic training (HVPT) (see e.g., Iverson & Evans, 2009; Lambacher et al., 2005; Nishi & Kewley-Port, 2007, 2008). HVPT is based on the assumption that training learners to perceive sounds that are produced by multiple speakers yields better results than listening training that only includes speech produced by one speaker (Thomson, 2012).

At the end of every production-based iCPR unit, the learners are prompted to upload a recording of their own pronunciation to a dropbox on the school's course management system, so that the teacher can check for homework completion. It is very crucial to point out, however, that the teachers do not need to listen to these recordings and provide feedback. Rather, the learners improve their pronunciation simply by working through the iCPR units, listening to native speaker models, and receiving automated feedback in the binary choice perception tasks. The teachers can, of course, choose to provide feedback, and it is possible that doing so will even increase the learning benefits for the students. Yet, teachers do not always have time to provide individual feedback; thus, it is important for teachers to know that this method has been shown to be beneficial to learners even in the absence of feedback (Martin, 2015; Tanner & Landon, 2009).

Perception Units

As stated above, the first day of pronunciation training always starts with an iCPR unit targeting perception (i.e., listening) training. These perception units include two types of listening exercises: an accentedness detection task and a sound discrimination task. The accentedness detection task uses native-speaker recordings of German words, each of which has a sound that is problematic for American English speakers learning German. The goal is to examine the learners' ability to discern native from accented productions of problematic L2 speech sounds in words that contain these sounds. To this end, ten new vocabulary items are presented to the students one-by-one on a Microsoft *PowerPoint* slide. On each slide, the students see the written version of a German word and can then listen to two recordings for the same word. One recording is always spoken with Standard German pronunciation, while the other recordings is intentionally rendered with an American English accent of the German word. The learners can listen to the recordings as often as they wish and receive the solution on the following slide (see Figure 1).

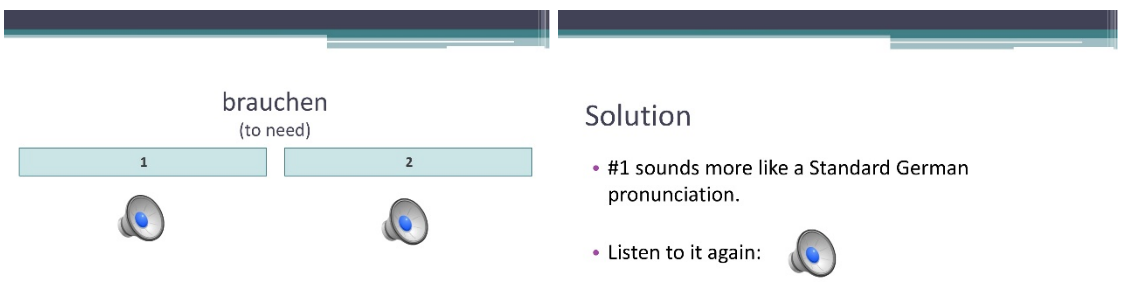


Figure 1. Sample slides for accentedness detection task in perception iCPR units.

The sound discrimination task then focuses on training the perception of problematic L2 speech sounds in contrast to similar sounds, which could easily be perceived as the targeted sound by an L2 speaker who is unfamiliar with the phonetic inventory of the L2. For example, when the focus is learning the German sound for the letter 'ü' (/y:/), the targeted sound would be contrasted with German 'i' (/i:/) and German 'u' (/u:/). For this exercise, the L2 speech sounds are embedded in nonsense words, that is invented words that are actually meaningless but have a sequence of sounds that is permissible in German. Nonsense words were chosen rather than real words so that listeners would focus entirely on the sound rather than the meaning of the word. Thus, based on the example above, learners would hear the targeted nonsense word 'püt' (/py:t/) contrasted with similar nonsense words like 'pit' (/pi:t/) and 'put' (/pu:t/). Just as in the first training block, students are presented with the nonsense words one by one on a Microsoft *PowerPoint* slide. This time, however, not the entire nonsense word is spelled out but only the orthographic correspondence of the targeted L2 speech sound (e.g. the letter 'ü' for the German sound /y:/). Again, there are two recordings given on each slide that students can listen to as often as they wish before skipping to the next slide on which they receive feedback about the correct answer (see Figure 2).

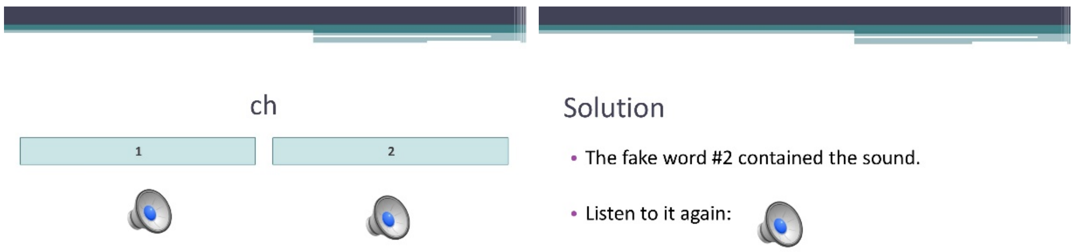


Figure 2. Sample slides for sound discrimination task in perception iCPR units.

Production Units

A day of listening activities is usually followed by two days of iCPR units focusing on production training. All production training starts with some explicit instruction delivered in the Microsoft *PowerPoint* presentation. When a new sound is introduced, for example, this means that learners read through some slides explaining what environments the sound is used in and how to configure the speech apparatus in order to produce the sound. This is followed by active practice. First, the learners repeat just the sound after a native speaker recording, then they progress to repeating entire words containing this sound. The practice phase consists of 10-15 individual words and the learners can listen to the native speaker recordings of each word as many times as they wish to practice. At the end of every unit, there is a *Practice Review* slide that contains a list of all words that were practiced in this iCPR unit and their corresponding recordings. This slide allows the learners to practice all words once more. The learners are then prompted to open recording software (e.g., *Audacity,* see next section) and to read all previously practiced words out loud (see Figure 3). The students upload these recordings to a dropbox on the school's course management system, so that the teacher can check for homework completion. I want to stress once more, however, that the teachers do not need to listen to these recordings and provide feedback.

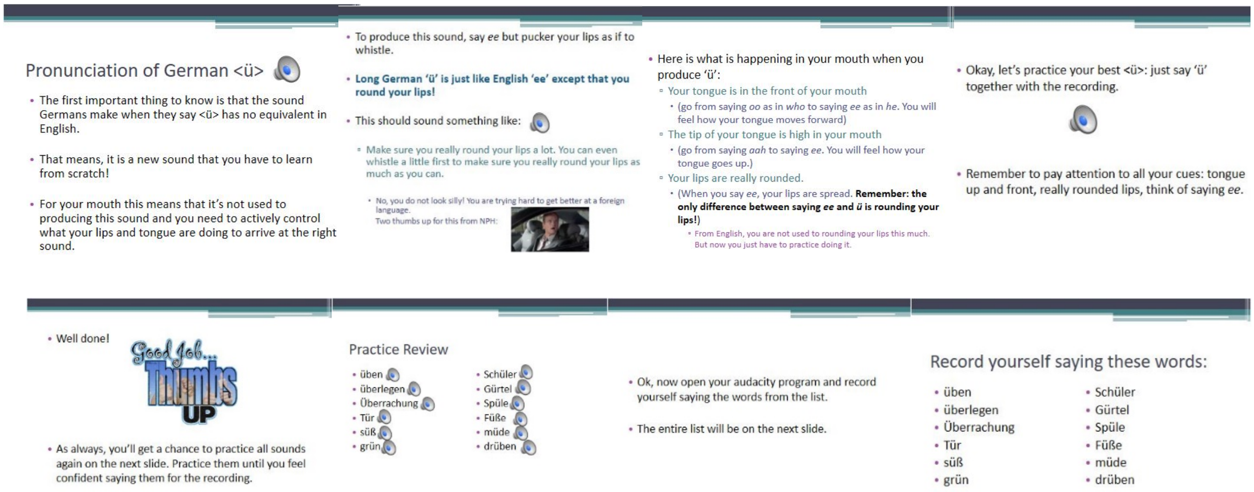


Figure 3. Sample slides for production iCPR units.

Application Tips

A complete set of 30 iCPR units (i.e. ten weeks of pronunciation instruction) for novice learners of L2 German can be accessed here: https://www.dropbox.com/sh/e3zolu9ffsr9x31/AABqiQKgxH1Xx8P2b56DkkYOa?dl=0.

So far, iCPR units have only been designed for German, but they can easily be adapted for any other second language. I would recommend starting the design process by coming up with a list of L2 sounds or prosodic features, such as stress placement, that have been shown to impede the intelligibility of your learners. The next step would be to compile a list of words that contain the problematic sounds. This can easily be done by going through the vocabulary pages of the learners' textbook. These will constitute the word inventory that you include for practice in the iCPR units. The next step is to create a new *PowerPoint* presentation or to simply replace the words in the *PowerPoin*t files that can be downloaded at the link provided above with your own words. It is very important that you provide a recording for each of the practice items. I recommend using *PowerPoint*'s built-in recording feature. The recordings created with this feature are of good quality, very easy to record and insert on a slide, and, most importantly, many recordings can be embedded without increasing the overall file size too much. When embedding recordings that were recorded with another software or even a phone, the file size can quickly become too large, which makes it less convenient for your students to download the files onto their personal computers. As mentioned above, ideally, you would alternate between multiple native speakers when creating these recordings to account for findings on the value of HVPT. If you do not have access to several native speakers, however, this should not discourage you from creating iCPR units. Martin (2015) has found the method to be beneficial to first semester learners, even when all recordings in the iCPR units employed in this study were recorded by a single native speaker only. It might even be beneficial when having advanced L2 speakers serve as language models for the recordings; however, the use of nonnative pronunciation models in iCPR units has not been tested empirically.

As part of the production iCPR units, learners are prompted to record themselves and to upload their recordings to a dropbox on the school's course management system. The goal behind this step is not to provide individual feedback to each learner, but rather to ensure that the learners complete the pronunciation homework exercises and that the teacher gets an overall impression of the learners' progress. For these recordings, I recommend *Audacity* recording software to my students (www.audacityteam.org). Audacity is a freeware program and very easy to use, and it is thus an easy solution for the learners.

Once the iCPR units for any language are created, they can easily be shared with colleagues at one's own school or at other schools. Just as with the creation of any effective teaching material, it certainly takes some time to create these units, but their design allows for them to be easily re-used and shared, which makes up for the time invested in their assembly. The design of three units per pronunciation topic further allows their flexible use as add-ons to an established curriculum. That is, if a teacher finds his/her students struggling only with some of the targeted sounds, s/he can always choose only to assign the problem-specific iCPR units as homework.

Distance Language Learning Environments

There has been a growth of learning environments over the past two decades, such as the development of online universities and hybrid courses, that is, on-campus classes with major online components. In light of these developments, it is important not only to think of language instruction in traditional face-to-face environments, but also in distance learning environments. Given that research on oral proficiency development, and particularly on pronunciation skills, in these new learning environments is still scarce (but see Blake, 2008; Blake, Wilson, Cetto, & Pardo-Ballester, 2008; Deutschmann, Panichi, & Molka-Danielsen, 2009; Isenberg, 2010), we currently do not know much about how to help distance learners improve their pronunciation skills in an L2. iCPR units, however, have great potential to be beneficial to online learners in this regard, as their design allows them to be integrated as an easy add-on in a distance learning environment.

CONCLUSION

Assigning pronunciation training as homework exercises by using iCPR units gives pronunciation training a place in every language curriculum without taking up valuable in-class time or singling out students in front of the class. The units employ easily accessible software, and once they are designed, they can be used by any teacher requiring no extra work or time commitment. Finally, iCPR units can also be used by teachers who do not feel adequately prepared to teach pronunciation in-class, allowing these teachers to give their students the additional benefit of pronunciation training without having to actively teach pronunciation in the classroom.

ABOUT THE AUTHOR

Ines A. Martin is a Ph.D. candidate in German Applied Linguistics and Language Science at the Pennsylvania State University. She received a teaching degree in German, French, and English as a foreign language from the University of Marburg in Germany and has been teaching all levels of German at Penn State since 2013. Her research focuses on L2 phonology and the acquisition of German pronunciation in face-to-face and distance learning environments. For her current dissertation research, she investigates new methods of pronunciation training that can easily be implemented in classrooms even by instructors with little or no background in teaching pronunciation. Email: martin@psu.edu

REFERENCES

Blake, R. (2008). Distance learning for second and foreign language teaching. In N. Van Deusen-Scholl & N.H. Hornberger (Eds.), *Encyclopedia of Language and Education: Second and Foreign Language Education* (pp. 365-376). Heidelberg: Springer.

Blake, R., Wilson, N.L., Cetto, M., & Pardo-Ballester, C. (2008). Measuring oral proficiency in distance, face-to-face, and blended classrooms. *Language Learning & Technology, 12*(3), 114-127.

Breitkreutz, J.A., Derwing, T.M., & Rossiter, M.J. (2001). Pronunciation teaching practices in Canada. *TESL Canada Journal, 19*(1). 51-61.

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

Deutschmann, M., Panichi, L., & Molka-Danielsen, J. (2009). Designing oral participation in Second Life – a comparative study of two language proficiency courses. *ReCALL, 21*(2), 206-226.

Derwing, T.M. (2013). Pronunciation instruction. In C.A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics* (pp. 1-9). Chichester, UK: Wiley-Blackwell.

Flege, J.E. (1995). Second language speech learning: Theory, findings and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in Cross-Language Research* (pp. 233-272). Baltimore, MD: York Press.

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

Iverson, P., & Evans, B.G. (2009). Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers. *The Journal of the Acoustical Society of America, 126*(2), 866-877.

Isenberg, N.A. (2010). *A comparative study of developmental outcomes in web-based and classroom-based German language education at the post-secondary level: Vocabulary, grammar, language processing, and oral proficiency development* (Unpublished doctoral dissertation). The Pennsylvania State University, University Park, PA.

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.

Martin, I.A. (2015). An easy way of teaching pronunciation in elementary level foreign language classes: The method of iCPRs. *Pennsylvania Language Forum*, *85*(1), p. 40-45.

Nishi, K., & Kewley-Port, D. (2007). Training Japanese listeners to perceive American English vowels: Influence of training sets. *Journal of Speech, Language, and Hearing Research, 50*, 1496-1509.

Nishi, K., & Kewley-Port, D. (2008). Nonnative speech perception training using vowel subsets: Effects of vowels in sets and order of training. *Journal of Speech, Language, and Hearing Research, 51*, 1480-1493.

O'Brien, M.G. (2004). Pronunciation matters. *Die Unterrichtspraxis, 27*(1), 1-9.

Tanner, M.W., & Landon, M.M. (2009). The effects of computer-assisted pronunciation readings on ESL learners' use of pausing, stress, intonation, and overall comprehensibility. *Language Learning & Technology, 13*(3), 51-65.

Thomson, R.I. (2011). Computer assisted pronunciation training: Targeting second language vowel perception improves pronunciation. *CALICO Journal, 28*(3), 744-765.

Thomson, R.I. (2012). Improving L2 listeners' perception of English vowels: A computer-mediated approach. *Language Learning, 62*(4), 1231-1258.

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

Zetterholm, E. (2017) Teaching the pronunciation of Swedish exotic vowels. 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. 222-227). Ames, IA: Iowa State University.

TEACHING TIP

TEACHING THE PRONUNCIATION OF SWEDISH EXOTIC VOWELS

Elisabeth Zetterholm, Stockholm University

INTRODUCTION

Learning a new language as an adult is often a hard task, particularly the pronunciation of new speech sounds. Earlier research has shown (e.g., Bannert, 1990; Zetterholm & Tronnier, 2017) that second language learners of Swedish have difficulties learning to pronounce some Swedish vowels, especially the front rounded vowels, and distinguishing them from each other and from unrounded vowels. For intelligible speech, it is of importance that there is an audible distinction in the pronunciation of the different rounded vowels. Otherwise a native listener might misunderstand the speaker referring to the number of minimal pairs with the vowel as the distinctive feature. In comparison with languages around the world, the Swedish vowel inventory is relatively large (Ladefoged, 2005), and some are exotic from a worldwide perspective. These vowels are one of the most difficult pronunciation features of Swedish pronunciation for second language learners to master.

The aim of this paper is to provide information about the vowels of Swedish with a focus on rounded vowels that are challenging for L2 learners of Swedish and to present some techniques to facilitate the accurate pronunciation of the exotic vowels.

Background

Swedish is a Germanic language closely related to the other two Scandinavian languages, Danish and Norwegian. It is not only the national language in Sweden (with approximately 9 million speakers) but also one of the national languages in Finland where 6 % of the population are Swedish-speakers. There are several regional dialects in Sweden that differ concerning the pronunciation of phonemes as well as the prosody, but the orthography is based on standard Swedish. The description in this paper follows the standard Swedish variety. However, second language learners of Swedish often learn the spoken variety in the area where they live and where they study Swedish.

What the teacher needs to know about Swedish vowels

There are nine basic vowel phonemes in Swedish (Bruce 2010; Riad 2014). The long vowels are phonemes and each phoneme has a short variant as an allophone, making in total 18 different vowel sounds (see Table 1). One could suggest that the vowels should be counted as 18 different phonemes. We follow Bruce (2010, p. 111) in this paper, who claims that there should be nine distinctive vowel phonemes and nine allophones.

There is not always a one-to-one correspondence between pronunciation and orthography in Swedish. However, it is a more transparent correspondence compared to the relationship between orthography and pronunciation in English (Katz & Frost, 1992). For example, the short vowel [e] is mostly realized as [ɛ] even though there is both an individual and dialectal variation. The short vowels [ø] and [ɛ] are pronounced differently depending on the following consonant as well as dialectal variation. The more open allophones [œ] and [æ] respectively are commonly used before an /r/. This is shown in words like *höna* [hø:na] (hen) and *höra* [hœ:ɾa] (hear) and *häl* [hɛ:l] (heel) and *här* [hæ:ɾ] (here). The long /ɑ:/ is slightly rounded, and not as unrounded as the short [a]. The letter <o> can be pronounced like in *son* [so:n] (son) – *sol* [su:l] (sun) – *som* [sɔ:m] (as/relative pronoun who/which) – Olle [ʊl:ɛ] (a male name).

Table 1

The Swedish vowels, IPA and orthography

	Long vowels, phonemes		Short vowels, allophones		Orthography	
	unrounded	rounded	unrounded	rounded		
Front	i:	y:	ɪ	ʏ	i	y
	e:	ø:	e/ɛ	ø/œ	e	ö
	ɛ:		ɛ/æ		ä	
			a		a	
Central		ʉ:		ɵ		u
Back		u:		ʊ/ɔ		o
		o:		ɔ		å
		ɑ:			a	

Exotic Swedish vowels

The two Swedish front rounded vowels /y:/ and /ø:/ and the central rounded vowel /ʉ:/ are quite uncommon among languages around the world. Referring to the theory of markedness (Eckman, 2008) these vowels could be described as marked. The three rounded back vowels /u:/, /o:/ and /ɑ:/ occur in many languages, but can be pronounced with great variation both by L1 and L2 speakers of Swedish.

Front unrounded vowels

As seen in Table 1, there are three front unrounded phonemes but six short allophones. The difference between them is mainly the degree of opening of the lower mandible. Notice that the short unrounded [a] is a front vowel and the long half-rounded /ɑ:/ is a back vowel. The difference between the four opening degrees is of importance for the distinction between the vowels. The highest, and most closed vowel is /i:/ followed by /e:/, /ɛ:/ and [a], when lowering the jaw. Some examples:

lika [li:ka] (alike) – *leka* [le:ka] (play) – *läka* [lɛ:ka] (heal)

rita [ri:ta] (draw) – *reta* [re:ta] (tease/make fun of) – *räta* [rɛ:ta] (straighten) – *ratta* [rat:a] (to drive a car with the steering-wheel), but *rata* [rɑ:ta] (reject) with a long back vowel.

Front and central rounded vowels

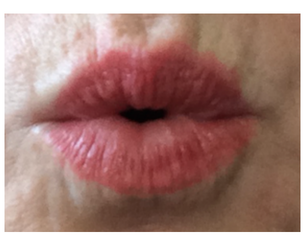
The two front rounded vowels /y:/ and /ø:/ have almost the same articulation as /i:/ and /e:/ respectively, except for the rounded lips. When the learner is aware of the different opening degrees of the jaw he/she can articulate the unrounded vowels and just make the lips rounded. The lips have to be outrounded for the two vowels /y:/, see Figure 1, and /ø:/ and the tongue close to the teeth in the lower mandible. When producing the vowel /y:/ some would say that the learner might think of the protrusion used when trying to kiss someone or sucking the thumb.

The least common vowel among languages worldwide is the Swedish /ʉ:/ in words like *hus* [hʉ:s] (house). To achieve a pronunciation that is separate and different from the back rounded vowel /u:/ the learner has to shape the lips like when whistling i.e, see Figure 2. A rather tense upper lip but not too outrounded lips like the pronunciation of /y:/, but the tongue at the same position close to the teeth in the lower mandible. However, the opening of the jaw should be at almost the same degree as for /i:/ and /y:/ respectively.


Figures 1 and 2. Lip protrusion when producing /y:/ (to the left) and /ʉ:/ (to the right).

Back rounded vowels

The Swedish back rounded vowel phonemes are /u: o: ɑ:/. The back rounded vowel /u:/ is pronounced with inrounded lips and /o:/ has outrounded lips but closer than /ø:/. When pronouncing the /o:/ the lips can be rounded as when blowing a candle carefully, see Figure 4. The long vowel /ɑ:/ is slightly rounded with outrounded lips, see Figure 3. Except for the rounding, the opening degree of the jaw is of importance. It is observed that second language learners have difficulties with the distinction between front and back or more central vowels, such as /ø:/ and /o:/ as well as /ɛ:/ and /ɑ:/. There is a great distinction concerning both the meaning and the pronunciation of the minimal pairs: *lön* [lø:n] (salary) and *lån* [lo:n] (loan) as well as *väl* [vɛ:l] (well) and *val* [vɑ:l] (whale or election).

Figures 3 and 4. Lip protrusion when producing /ɑ:/ (to the left) and /o:/ (to the right).

General approach

To be aware of and feel the articulation in the mouth – the opening degree of the jaw and the different types of lip rounding is a good starting point. Hyper articulation and a mirror are helpful. A native speaker of Swedish does not always exaggerate the lip rounding but still changes the lip protrusion for inrounded and outrounded vowels. However, for a second language learner it might be hard to perceive and pronounce the subtle, but important, differences in articulation. Therefore, perception is crucial for the production of the Swedish vowels, at least for some vowels that are quite uncommon among languages worldwide. Imitation and shadowing are useful tools for teaching. Recordings and a careful and critical listening of one's own voice and pronunciation as well as other second language learners are often fruitful ways to consciousness.

Proposed teaching tips

Learners of Swedish often think they do change the protrusion of the lips and maybe they do, but only to some degree. Most learners do not produce a clear difference between the vowels. Therefore, a mirror is very useful. To be aware of the articulation, the position of the jaw and the lip-rounding you can use these teaching tips:

1. Say /iː eː ɛː/ and [a] (the quality of the short [a]) and focus on the opening of the jaw.
2. Say /iː yː/ and focus on the difference between the unrounded and the rounded vowel.
3. Say /eː øː/ and focus on the difference between the unrounded and the rounded vowel.
4. Say /yː ʉː uː/ and focus on the protrusion of the lips.
5. To be aware of the difference between a long slightly rounded /ɑː/ and the short unrounded [a] you may listen to the difference and imitate word pairs. Here some examples of words with one syllable: *tal* [tɑːl] (speech or number) – *tall* [talː] (pine); *mat* [mɑːt] (food) – *matt* [matː] (faint); words with two syllables and the stress on the first syllable: *haka* [hɑːka] (chin) – *hacka* [hakːa] (chop), *baka* [bɑːka] (bake) – *backa* [bakːa] (reverse). In these examples, the orthography is a clue for the duration of the vowel as well as the consonant.

Look in a mirror to be aware of how much you have to open the jaw and round the lips to produce a clear difference between the vowels. In order to exaggerate the lip protrusion you can put a pencil above the upper lip, keep it there when pronouncing an /yː/ for outrounded lips, as in Figure 1. Change the lip protrusion of the upper lip for the pronunciation of an /ʉː/ and you will drop the pencil, see Figure 2. For the pronunciation of /uː/ the lips are more inrounded. Now you will probably change the tongue position and lower the jaw as well. This vowel, /uː/, is a phoneme in many languages and therefore it might not be a pronunciation problem for second language learners of Swedish in general.

Practice minimal pairs to hear the differences and be aware of the meanings that could lead to misunderstandings. A picture of a vowel chart is helpful, especially for comparison with vowels in the learners' first language. This is often helpful in pronunciation teaching in many languages for awareness of articulation and production of sounds that might be similar or different.

Minimal pairs with rounded and unrounded vowels

fira [fiːɾa] (celebrate) – *fyra* [fyːɾa] (four)

bita [biːta] (bite) – *byta* [byːta] (change) – *bota* [buːta] (cure)

len [leːn] (smooth) – *lön* [løːn] (salary) – *län* [lɛːn] (county) – *lån* [loːn] (loan)

har [hɑːɾ] (have) – *hör* [hœːɾ] (hear) – *här* [hæːɾ] (here) – *hår* [hoːɾ] (hair) – *hur* [hʉːɾ] (how) – *hyr* [hyːɾ] (rent)

läsa [lɛːsa] (read) – *lösa* [løːsa] (solve) – *låsa* [loːsa] (lock) – *lysa* [lyːsa] (shine)

Minimal pairs and contrasts with the vowels /ʉː/ and [ɵ]

utan [ʉːtan] (without) – *ytan* [yːtan] (the surface)

sur [sʉːɾ] (sore or soggy) – syr [syːɾ] (sewing)

sila [siːla] (drain) – *sula* [sʉːla] (sole) – *sola* [suːla] (sunbathe)

lycka [lykːa] (happiness) - *lucka* [lɵkːa] (gap)

flyga [flyːga] (to fly) – *fluga* [flʉːga] (fly, an insect)

ryta [ryːta] (roar) – *ruta* [rʉːta] (square)

fil [fiːl] (file) – *ful* [fʉːl] (ugly)

full [fɵlː] (drunk) – *fall* [falː] (fall) – *fäll* [fɛlː] (fleece) – *fåll* [fɔlː] (hem) – *föll* [følː] (fell)

CONCLUDING REMARKS

The pronunciation and the contrast between unrounded/rounded and front/back vowels are not difficult for a native speaker of Swedish. The articulation and protrusion is not exaggerated, but still, a distinctive lip-rounding is of importance. However, for a second language learner of Swedish, one of the most difficult pronunciation features among the vowels seems to be the front and the central rounded vowels. Many learners, regardless of their first language, do not round the lips to a degree that makes an audible distinction between the unrounded and the rounded vowels. Exceptions are found in recordings with L1-speakers of Albanian and Finnish (Zetterholm & Tronnier 2017) where the distinction is clear. The short front vowel [y] is often replaced with an [i], *cykla* [sykla] (to bike) – **sikla* [sɪkla] (the symbol * means a non-word in Swedish). The central /ʉː/ is often replaced with the back vowel /uː/, *hus* [hʉːs] (house) – *hos* [huːs] (among) and the short variant in *buss* [bʉsː] (bus) is replaced by **boss* [busː] (a non-word). Both /iː/ and /uː/ occur in other languages and that might be one explanation for the substitute and mispronunciation in Swedish. Another observation concerning the rounded vowels is that the learner has to be aware of the differences between outrounded and inrounded lips as well as the degree of the rounding. In the Swedish movie *Se upp för dårarna* (Nutley & Bergström 2007) (English title *Mind the Gap*), this was noted and pointed out in conversations between a father and his daughter, both second language learners of Swedish. The father has a heavy accent and has a job as an underground train driver (although he is a heart surgeon from Turkey). The driver announces that the door will be closed and should say *Se upp för dörrarna* [seː ɵpː fœːɾ dœɾːaŋa] (mind the doors). Unfortunately, with his accent and his lip-rounding he says *Se upp för dårarna* [seː ɵpː fœːɾ doːɾaŋa] (mind the fools). In this context you will understand his message about the doors, but it could be misunderstood as another kind of warning as well. There is also a quantity distinction between the two highlighted vowels [œ] and [oː]. This mispronunciation of rounded vowels seems to be a challenge to overcome for many second language learners of Swedish. However, there are some clues and teaching tips that might be helpful. As a starting point, the use of a mirror to be aware of the protrusion of the lips is helpful.

ABOUT THE AUTHOR

Elisabeth Zetterholm is a Senior Lecturer and Associate Professor in Swedish as a second language at the Department of Language Education, Stockholm University, Sweden. Her research interests are second language learning, literacy, multilingualism, phonology and pronunciation, both from the teachers' and the learners' viewpoints. When teaching, most of her students are prospective teachers of Swedish as a second language.

REFERENCES

Bannert, R. (1990). *På väg mot svenskt uttal*. Lund: Studentlitteratur.

Bruce, G. (2010). *Vår fonetiska geografi*. Lund: Studentlitteratur.

Eckman, F. R. (2008). Typological markedness and second language phonology. In J. G. Hansen Edwards & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 95-115). Amsterdam: John Benjamins.

Katz, L., & Frost, R. (1992). The reading process is different for different orthographies: The orthographic depth hypothesis. *Haskins Laboratories Status Report on Speech Research*, SR-111/112, 147-160.

Ladefoged, P. (2005). *Vowels and consonants*. Blackwell Publishing Ltd.

Nutley, C. (Producer) & Bergström, H. (Director) (2007). Mind the Gap (English title) *Se upp för dårarna* (Swedish title).

http://www.sfi.se/en-gb/Swedish-film-database/Item/?type=MOVIE&itemid=62474

Riad, T. (2014). *The phonology of Swedish*. Oxford University Press.

Zetterholm, E., & Tronnier, M. (2017). *Perspektiv på svenskt uttal. Fonologi, brytning och didaktik*. Lund: Studentlitteratur.

Chen, M. (2017). Saundz (Review). 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. 228-235). Ames, IA: Iowa State University.

SOFTWARE REVIEW

Saundz

Mo Chen, Iowa State University

The *Saundz* curriculum-based education software program was developed by the *Saundz* Company in 2012. Users can access the *Saundz* website on their computers or download the *Saundz* app to their smartphones and tablets via Apple's iOS or the Android operating system. The cost for access to the full course on the website and through the app is $19.99 USD. *Saundz* is intended to offer computer-assisted pronunciation training (CAPT) to English-as-a-foreign-language (EFL) students. The developers of *Saundz* promise that the software will help non-native English speakers to "1) learn in a step-by-step manner each of 40 sounds of American English, 2) practice their pronunciation at any time via the internet, and 3) quickly improve their pronunciation skills and reduce their accent" (Saundz, 2016). This review will focus on the *Saundz* app. The review will begin with an overview of the main features of the app, and it will then evaluate the app using the CALL evaluation framework developed by Chapelle (2001).

OVERVIEW OF MOBILE APPLICATION

This section will discuss the four main features of the *Saundz* app: 1) virtual pronunciation instructors and teaching assistants, 2) curriculum-based lessons, 3) customized curriculum and 4) the grading and badge systems.

VIRTUAL PRONUNCIATION INSTRUCTOR AND TEACHING ASSISTANTS

One of the central features of *Saundz* is its computer-animated virtual pronunciation instructor, Simone, whom users meet when they select a lesson on the app. Simone's face, viewed head-on and in profile, demonstrates how to move the vocal organs in order to pronounce particular vowel and consonant sounds, as shown in Figure 1. Users can click on the "Play" button to watch animations and to listen to corresponding sounds as many times as they choose. In addition to visual and audio cues, a detailed written description of how to pronounce the sounds is provided (see Figure 1).

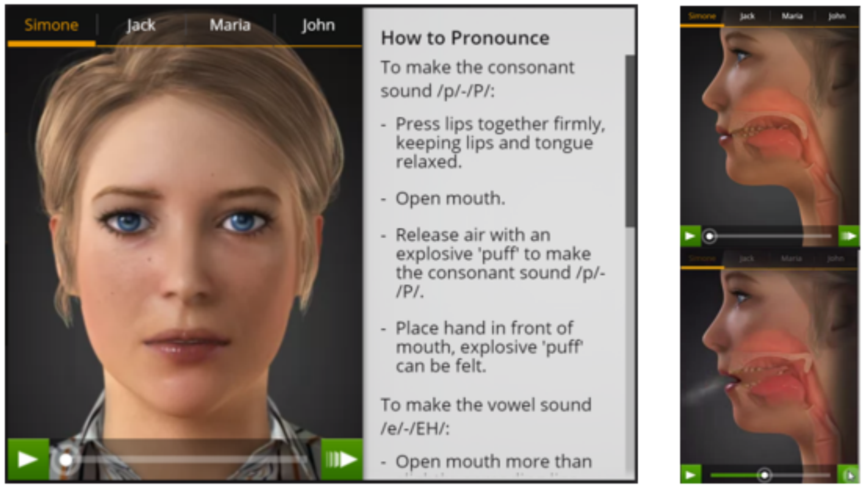


Figure 1. Simone's demonstrations for making the consonant /p/ sound

In addition to Simone's voice, users can choose to listen to the voices of three professional English teaching assistants: Jack, Maria or John. Jack is an adult male, while Maria and John are teenagers (see Figure 2). Users can choose the teaching assistants based on their age and gender.

Figure 2. The icons of the teaching assistants Jack, Maria and John

CURRICULUM-BASED LESSONS

The full course of *Saundz* includes 38 chapters with 163 lessons to teach the pronunciation of 16 vowel and 24 consonant sounds in English. In each lesson, students first learn from their virtual pronunciation instructor, Simone, how to produce a target sound (e.g. /e/). Next, they can listen to the target sounds as they appear in different words (e.g. pet, peck and pep). In total, the system includes more than 400 words to teach students how to produce various sounds. The definitions of each word, along with pictures and sample sentences, are provided in a dictionary in the left sidebar of the window (see Figure 3).

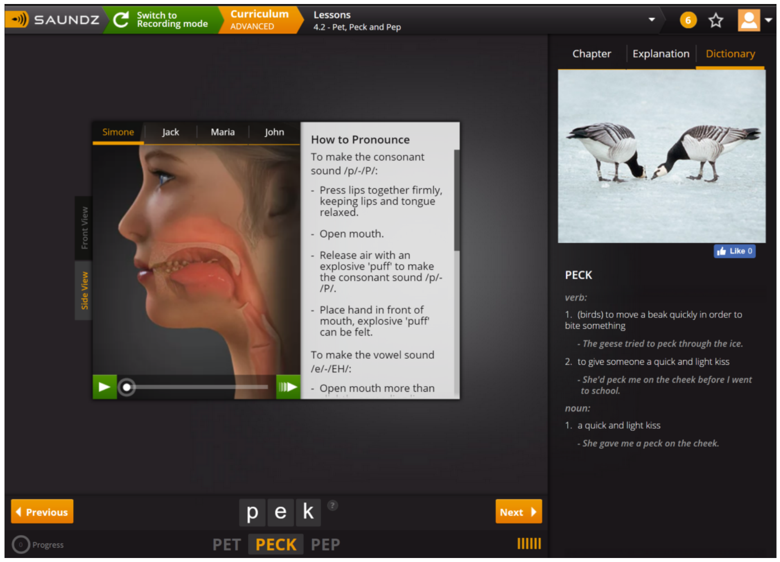


Figure 3. A screenshot of the lesson on /e/ with the words "pet," "peck," and "pep"

Minimal pair exercises are traditional exercises that are widely used in language training to encourage students to perceive discrete sounds (Munro & Derwing, 2006). After learning the phonetic knowledge of a particular sound, *Saundz* users can listen to target word pairs (e.g. pep vs pit, peck vs pack, tech vs tack) multiple times to notice the subtle differences between two words. Furthermore, they can read those word pairs aloud and compare their productions with Simone's or those of the teaching assistants in interactive recording exercises. Users record their production of a target word as many times as they wish and select the three best recordings. They then upload those three audio files to the system to compare them to the sounds produced by the virtual instructors (see Figure 4). Users are expected to improve their perceptions and productions of the target sounds with minimal pair listening and recording activities.

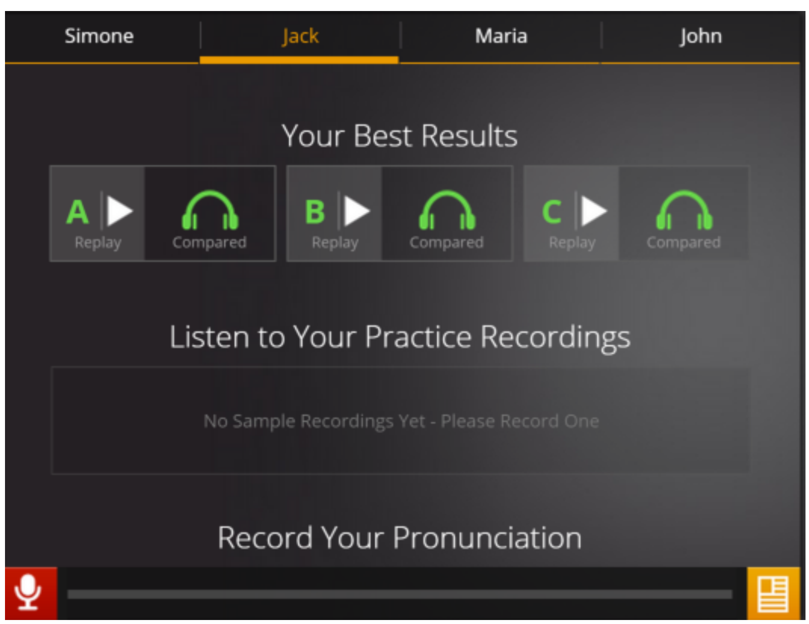


Figure 4. A screenshot of the recording mode in *Saundz*

CUSTOMIZED CURRICULUM

As mentioned above, *Saundz* has 163 lessons focusing on 16 vowel and 24 consonant sounds in English. Based on learners' first languages, the course recommends target lessons that address problematic sounds for native speakers of certain languages. For example, the course recommends 134 lessons focusing on 28 sounds to learners whose native language is Chinese. As shown in Figure 5, Chinese students are encouraged to make use of lessons about the distinctions between consonants such as /r/ and /l/, the length contrast between the vowel sounds /i/ and /ɪ/ and the pronunciation of problematic sounds like /v/. By comparison, the course suggests that Japanese speakers review 94 lessons focusing on 24 sounds. These lessons address the differences between /f/ and /v/, the pronunciation of the diphthongs /aɪ/, /eɪ / and /ɔɪ / and the pronunciation of word-final consonants like /n/.

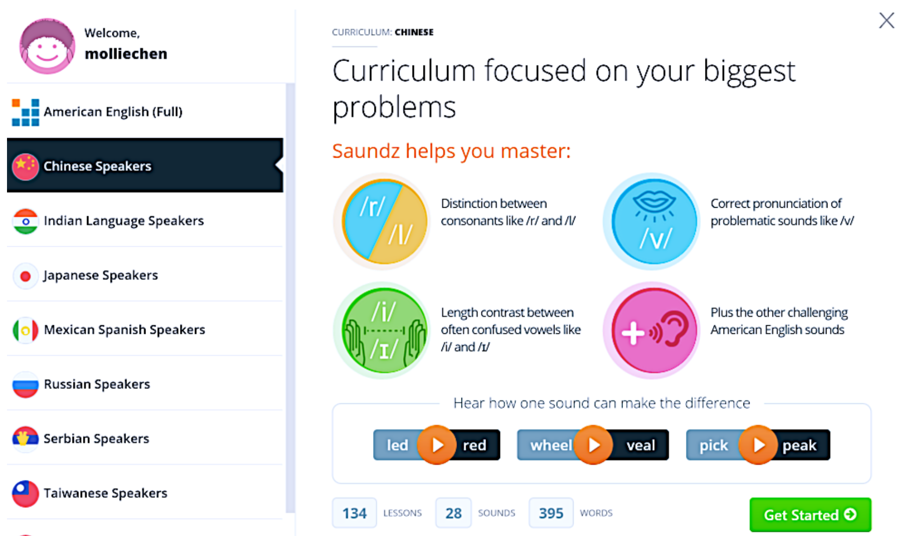


Figure 5. A screenshot of the front page of the Chinese curriculum in *Saundz*

THE GRADING AND BADGE SYSTEMS

In order to motivate students to continue interacting with the system, *Saundz* provides users with scores and digital badges as incentives. For example, after uploading recording files to *Saundz*, users receive scores for task completion, as shown in Figure 6.

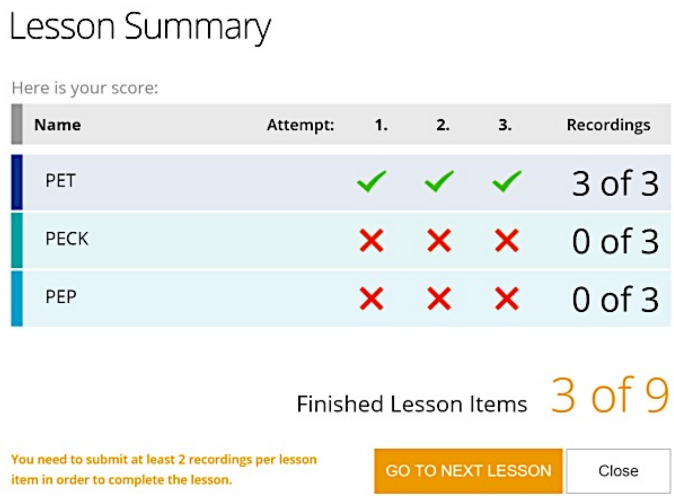


Figure 6. A screenshot of the scoring system in *Saundz*

The app also awards digital badges to students as they progress through the program. For example, when they finish a difficult lesson, such as a lesson on the vowel sound *schwa*, or achieve a goal, such as recording and uploading 50 words to *Saundz*, they can collect digital merit badges in the gallery as awards (see Figure 7).

Figure 7. A screenshot of the Merit Badges Gallery in *Saundz*

APPLICATION EVALUATION

This review implements Chapelle's framework for evaluating computer-assisted language learning (CALL) (2001). Based on this framework, I will assess the *Saundz* app from six perspectives: language learning potential, learner fit, meaning focus, authenticity, positive impact and practicality.

LANGUAGE LEARNING POTENTIAL

According to Chapelle (2001), language learning potential refers to the "degree of opportunity present for beneficial focus on form" (p.55). As Schmidt (2001) articulated, second language acquisition "is largely driven by what learners pay attention to and notice in target language input and what they understand the significance of noticed input to be" (p.4). In order to help students focus on target vowel and consonant sounds, *Saundz* uses animated virtual instructors and written descriptions. Moreover, the system also allows students to compare their own productions to the standard pronunciations through minimal pair listening and recording exercises. The comparisons enable students to notice gaps or mismatches between their productions and the model sounds.

Even though noticing is the first step in language learning, the noticing itself cannot ensure learning (Swain, 2005). One potential problem that I found in the comparison activities is that there is no corrective feedback provided to learners. As Sheen (2011) stated, corrective feedback can notify students of errors in their output and allow them to understand how to make changes. Through repeated practice with corrective feedback or exposure to model sounds, users are more likely to become aware of their own problems in pronunciation and to improve accordingly.

Another aspect of the app that can be improved concerns its grading and incentive systems. When users record and upload their production of the target words to the system, they receive scores. These scores are given based on students' completion of tasks. In other words, users receive full scores when uploading their recordings to the program, regardless of whether they produce the sounds correctly or incorrectly. As a result, the scores are a poor indicator of students' mastery of the knowledge. If users were able to receive scores based on the quality of the sounds they produce, the interactions with the learning system would be more meaningful and motivating.

LEARNER FIT

"Learner fit" refers to how much a program or a course benefits students with various characteristics (Chapelle, 2001). The developers of *Saundz* have tried to personalize the app by addressing different language backgrounds and proficiencies. First, *Saundz* recommends different chapters to non-native English learners with different language backgrounds. Given that students with the same L1 may have similar pronunciation problems, this is an effective way to address problematic sounds within target groups of users. As an English learner whose first language is Chinese, I found some lessons, such as those on the differences between the /i/ and /ɪ/ sounds, useful. Nevertheless, I noticed that some common pronunciation issues found among Chinese students are not addressed in the course. For instance, many Chinese students from southern provinces of China have difficulty differentiating the /n/ and /l/ sounds in their L1; those two sounds are also negatively transferred to their English pronunciation. *Saundz* fails to include a lesson explaining those two consonant sounds to Chinese users. Therefore, more chapters on confusing vowel and consonant errors should be developed and integrated into the existing lessons in *Saundz*.

A central claim made by the designers of *Saundz* is that the app "allows students regardless of their language proficiency, to learn American English pronunciation quickly" (Saundz, 2016). Admittedly, the three-dimensional technology and virtual instructors are very helpful; users with different language proficiencies can use the app's three-dimensional animation and written descriptions to understand how to produce target sounds. However, it is hard to say if users with intermediate or advanced language proficiency would benefit from the app's pronunciation exercises as much as beginners. Compared to beginner-level users who need to pay greater attention to sounds at the syllable- or word-level, users with higher language proficiency require more attention to suprasegmental features of utterances (e.g. intonation, stress, pitch, etc.) in particular contexts. In order to meet a variety of student needs, animations, exercises and tests that address pronunciation problems at the sentence and discourse levels ought to be integrated into the app.

MEANING FOCUS AND AUTHENTICITY

In Chapelle's framework (2001), "meaning focus" and "authenticity" refer to the connection between CALL activities and communication skills outside of the classroom. In terms of "meaning focus," the CALL exercises should not only direct users' attention to the pronunciation of the target sounds, but also allow them to understand the link between their productions and their intended meanings. What is more, the language use in the exercises should be authentic and represent the oral pronunciation features found in daily communication ("authenticity"). The exercises in *Saundz*, however, are neither meaning-focused nor authentic. First, the pronunciation training and practice address sounds and pronunciation problems entirely at the word level. None of the exercises are task-based or content-based. Second, the sounds produced by the virtual instructor and assistants are slower than those produced in real speech; users are not able to compare average-speed pronunciations to the ones in *Saundz*. The lack of meaning-focused exercises and authentic training materials may limit users' pragmatic knowledge development and understanding of how to apply what they learn in *Saundz* to real situations.

POSITIVE IMPACT AND PRACTICALITY

Saundz claims that "[i]n as little as 5 hours, from more than 40 available hours of instruction, ESL students using *Saundz* will improve their spoken American English with an overall reduction in accent." Developers of this software want to promote the positive impact of the app, emphasizing its practicality. However, this claim exaggerates the effects of the application. First, since accents appear not only at the word level, the improved pronunciation of discrete words as practiced through *Saundz* will not necessarily help users to achieve "an overall reduction in accent." Second, the appearance of an accent relates to various factors. Consequently, it is impossible to reduce accents within a short duration of time. The claims concerning the app's potential are therefore exaggerated.

In addition, *Saundz* (2016) provides testimonials from users from different countries:

> Excellent program! I've been trying to learn to speak correct for several years and *Saundz* finally helped me. (Jun—a learner from China)
>
> After I finished all the lessons in *Saundz*, everybody around me noticed my progress. (Galina—a learner from Serbia)

Significantly, all the evaluations provided by the developer come from insider perspectives (e.g. ESL/EFL learners and instructors). Evaluations from outsiders (e.g. applied linguists, program coordinators, etc.) are needed.

CONCLUSION

Overall, *Saundz* is a beneficial tool to help users improve their pronunciation at the word level. The virtual instructors and animations are useful to students who do not have access to native English speakers, and users can practice their pronunciation via mobile devices at any time and location. The app would benefit from the addition of visual and individualized feedback as well as contextualized pronunciation activities.

REFERENCES

Chapelle, C. (2001). *Computer applications in second language acquisition*. Cambridge University Press.

Munro, M. J., & Derwing, T. M. (2006). The functional load principle in ESL pronunciation instruction: An exploratory study. *System*, *34*(4), 520-531.

Saundz. (2016). American English Pronunciation Software. Retrieved from http://saundz.com/

Schmidt, R. (2001). Attention. In *Cognition and second language instruction, P. Robinson (Eds.),* Cambridge: Cambridge University Press, 3-32.

Sheen, Y. (2011). *Corrective feedback, individual differences and second language learning*. Springer.

Swain, M. (2005). The output hypothesis: Theory and research. *Handbook of research in second language teaching and learning*, *1*, 471-483.

Edalatishams, I. (2017). LeaP corpus (Review). 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. 236-240). Ames, IA: Iowa State University.

CORPUS REVIEW

LeaP Corpus

Idée Edalatishams, Iowa State University

INTRODUCTION

The LeaP corpus is a collection of speech by L2 learners of German and English, annotated mostly for phonetic features that contribute to prosody (Gut, 2014a). It was created as part of a larger project titled *Learning Prosody in a Foreign Language* aimed at describing learners' acquisition of prosody at both phonetic and phonological levels as well as the learner characteristics that affect the process of learning prosody (Gut, 2014b). The corpus was developed at the University of Bielefeld, Germany, from 2001 to 2003.

The Speakers

The LeaP corpus consists of speech from a total of 131 speakers varying in age, gender, L1, proficiency, age of first L2 exposure, length of L2 exposure, and non-linguistic factors such as motivation or musicality. Especially proficient L2 learners, or "superlearners" as Gut (2014a) calls them, were included among the speakers in order to provide data for exploring "ultimate phonological attainment" and the non-linguistic factors that contribute to native-like achievement (Gut, 2012). Some learners were recorded before and after receiving prosody training, so as to provide evidence for the effects of guided learning on pronunciation. Other learners were recorded before and after going abroad in order to provide information on the effects of unguided training on pronunciation. In addition to L2 learners, native speakers of English and German were also included to serve as control groups (Gut, 2014a).

Corpus Development

The speakers were recorded reading aloud a list of non-words and a short story, retelling the same story in their own words, and responding to informal interview questions. The resulting 12 hours of recordings were transcribed and annotated using manual and automatic methods and are available for free download in form of 359 annotated and time-stamped *Praat* TextGrids (Boersma & Weenink, 2016). The annotations include linguistic features of speech in 8 different tiers: phrase, word, syllable, segment, tone, pitch, part of speech, and lemmata. Additionally, the files are annotated for non-linguistic information—metadata—such as various speaker characteristics, date and place of the recording, and the language of the interview. After receiving intensive training, a total of six annotators carried out all the annotations and transcriptions. Inter-annotator and intra-annotator reliabilities were calculated and are reported to have differed considerably based on the complexity of each annotation task (Gut, 2012).

Corpus Use

The corpus is available for free download at https://sourceforge.net/ as well as for online access as part of the language archive at https://corpus1.mpi.nl/. The corpus is located on the side menu under "TLA corpora > donated corpora." Users can download the German and the English corpora separately. Each package is composed of all the sound files along with the annotations and metadata in *Praat* TextGrid and *XML* formats, resulting in 3 data files for each recording. The sound files can be played using any audio player application, but *Praat* is required for accessing the annotations along with the sound. In order for that, users need to open both the sound file and the *Praat* TextGrid file for each recording. After opening an annotated file in *Praat,* users can see 8 types of annotations at the phrase, word, syllable, segment, tone, pitch, part of speech, and lemma level. These tiers are not necessarily lined up in the same order, and some recordings do not include the last two tiers of annotation. Figure 1 illustrates a portion of an annotated file opened in *Praat* with all the 8 tiers of annotation for the sentence *I think it was very helpful for me*.

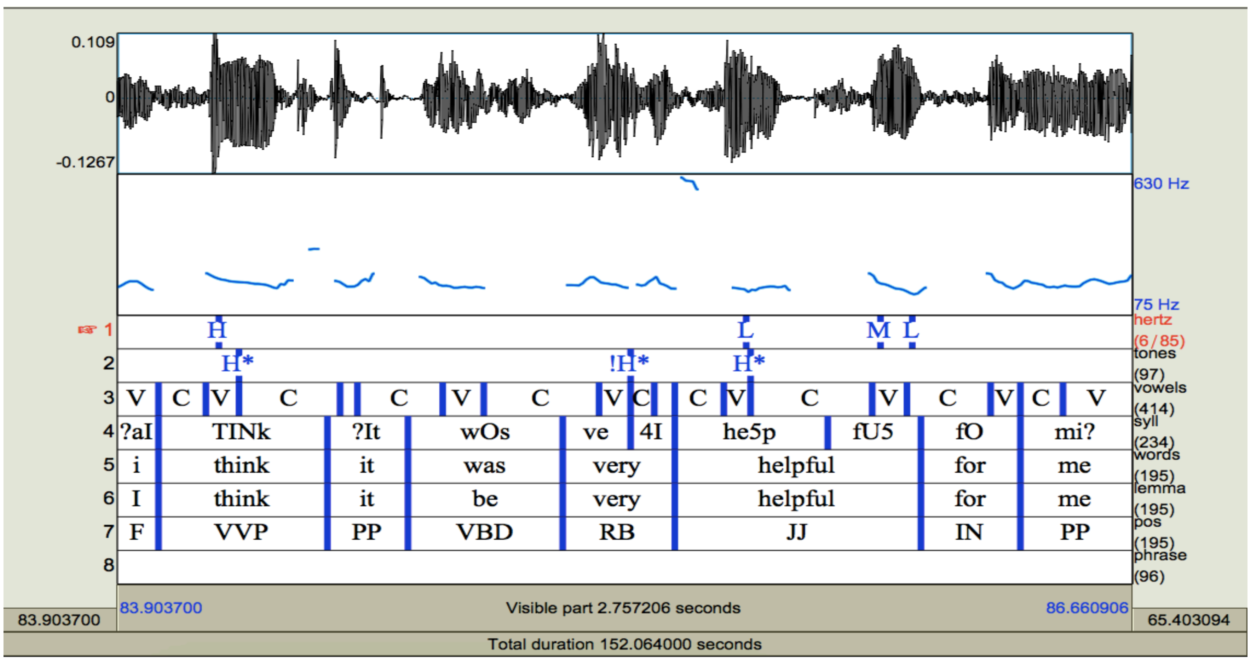


Figure 1. Sample Annotated Speech in the LeaP Corpus.

At the phrase level, in addition to the marked "quasi-intonation phrases," non-speech events such as laughter, noise, or breath are marked (Gut, 2012). The word tier includes the beginning and end of each word along with a manual transcription. At the syllable level, beginning and end of syllables are marked, while SAMPA symbols are used for a broad transcription of the syllable including only a few articulatory and coarticulatory phenomena such as aspiration, unreleased stops, and nasalization (Gut, 2012). At the segment level, vocalic intervals, consonantal intervals, and pauses are annotated. The tone tier includes annotations for pitch accents and boundary tones using a modified version of ToBI transcription system that accounts for phonetic—rather than phonological—realizations of tones (Gut, 2012). The pitch annotation tier includes markings for initial high pitch, final low pitch, pitch peaks, and pitch valleys. The two additional tiers have been automatically annotated and provide users with the part of speech and the lemma for each word.

In addition to manual search in the data files, Gut (2009) states that analysis of the corpus is also possible through *TASX* corpus browser. *TASX* is an XML-based data format specifically designed for this corpus. Browsing this type of data is possible using a set of scripts that can be accessed upon request sent to the corpus developer. According to Gut (2009), the user-friendly environment of *TASX* allows for searching and browsing in the data, running some statistical analyses, and converting to and from different file formats commonly used in phonetic research.

Besides browsing *TASX* data files, Gut (2012) states that the files can be converted to *XML*-annotated *NITE* format to be used within the *NXT* search tool, *NXT Search*, available at http://groups.inf.ed.ac.uk/nxt/index.shtml. *NXT* provides tools and libraries that enable "native representation, manipulation, query and analysis of multimedia language data" (Kilgour, 2017). This tool enables attribute tests as well as structural and temporal relations. For instance, searches can be done in words with a syllable containing a specific vowel (Slavianova, 2007) or for pitch accents on non-content words in non phrase-final position (Gut, 2014b). In the user manual, downloadable with the LeaP package, Gut (2014b) refers to the LeaP database as another tool for searching the corpus. This database, she states, enables easy generation of subcorpora. She also mentions that a user interface for online use of the corpus is under development.

Evaluation

In line with the purpose of the LeaP project, the LeaP corpus is expected to focus on learners' acquisition of prosody and learner characteristics that influence this process. Prosody is concerned with "parameters such as duration, intensity, and f_0 that contribute in various combinations to the production and perception of stress, rhythm and tempo, lexical tone, and intonation of an utterance" (Fletcher, 2012, p. 523). The LeaP corpus provides annotations for several prosodic features such as pitch measurements, pitch accents, and boundary tones. Parameters like fundamental frequency (f_0) and other vowel quality measures such as f_1, f_2, and f_3 can also be accessed through scripts that can be run on the corpus files. Opening the TextGrids along with the audio files in *Praat* will also enable users to access features such as vowel duration and energy level. Accounting for all these suprasegmental features, the corpus allows for extracting information about a whole variety of prosodic characteristics of learner language.

However, there exists an arguably significant problem with respect to the usability of the corpus. While the user manual introduces and provides links to three tools for semi-automatic searching and browsing of the corpus (*TASX* corpus browser, *NXT Search,* and the LeaP database), none of these tools seem to be publicly available. I personally contacted the corpus developer, Ulrike Gut, and she kindly shared the *TASX* browser scripts as well as scripts for converting *TASX* files into XML, required for using *NXT* search. Nevertheless, some expertise in programming in *Perl* language seems to be needed for working with these tools. Without such expertise, users are left with the rather inconvenient option of manually looking at the data, using the *Praat* TextGrid files along with the sound files. Needless to say, manually opening and reading each file in the corpus requires an extensive amount of time and does not provide a reliable approach to analyzing the data. As Rohlfing et al. (2006) have mentioned, readability of annotations through a limited number of tools is a drawback for multimodal annotation tools. In the case of this corpus, not only are the annotations not conveniently and accurately searchable, but also only *Praat* can be used for manual searching of the data.

Provided that the user has the necessary expertise to work with Perl scripts, the *NITE* tool, as one method to browse this corpus, would still face criticism. Slavianova (2007) has voiced concerns over the extensive amount of time the tool will require if all data files of the corpus are loaded into it. A corpus-wide query seems almost impossible as the *NITE* tool has not been designed to process large amounts of data in a reasonable amount of time. In response to these shortcomings, Slaviaona (2007) has developed the LeaP database as an alternative browsing tool for this corpus. However, the LeaP database is also not publicly available.

Apart from usability, reliability of the annotations of the LeaP corpus could also be improved. As Gut (2009) has reported, perfect inter-annotator and intra-annotator agreements were only achieved at the word, segment, and pitch tiers, whereas the syllable and tone tiers showed very low reliabilities. Gut attributes the low reliability for the syllable tier to the fact that the annotators had to carry out syllable segmentation and transcription simultaneously. This defect could be addressed by simply separating the two tasks at the syllable level into two different ones. Gut (2009) has further reported that the highest intra-rater reliability was achieved by only one of the six annotators, who had had previous experience with prosody annotation. She also associates the low reliability values for the inexperienced annotators to their lack of experience with and/or extensive training in annotation. While these weaknesses are acknowledged by the corpus creator, further improvements could be made in this regard by having experienced annotators check and recheck the annotations.

Despite all the aforementioned shortcomings and its relatively small size compared to most English learner corpora, the LeaP corpus offers rich annotations useful for a wide range of studies on L2 prosody. In fact, refuting the consideration of sample size as the most important factor in achieving representativeness in a corpus, Biber (1993) maintains that emphasis should instead be placed on the range of text types and linguistic distributions. The LeaP corpus achieves this goal to a sufficient degree by accounting for different speech types, diversity in L1s, and guided and unguided pronunciation training the speakers had received. Not only are these factors among the most frequently studied in phonetics research (Colantoni, Steele, & Escudero, 2015), but also considering the issue of practicality in spoken corpus development (Reppen, 2010; Adolphs & Knight, 2010), the amount of data and type of annotations in the LeaP corpus are adequate for it to be utilized in addressing many questions on the relationship between L2 prosody and different learner characteristics, linguistic contexts, or speech events. The LeaP corpus can be a starting point for filling what Colantoni, et al. (2015) refer to as the "theoretical lacuna" existing in L2 speech learning models with regard to prosody.

REFERENCES

Adolphs, S. & Knight, D. (2010). Building a spoken corpus. In M., McCarthy & A. O'Keeffe, *The Routledge Handbook of Corpus Linguistics*. New York, NY: Routledge.

Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing, 8*(4). 243-257.

Boersma, P. & Weenink, D. (2016). *Praat: doing phonetics by computer*. Retrieved from http://www.fon.hum.uva.nl/praat/

Colantoni, L., Steele, J., & Escudero, P. (2015). *Second language speech: Theory and practice.* Cambridge, UK: Cambridge University Press.

Fletcher, J. (2012). The prosody of speech: Timing and rhythm. In W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.) *Blackwell handbooks in linguistics: The handbook of phonetic sciences* (pp. 523-602). Hoboken, GB: Wiley-Blackwell.

Gut, U. (2009). *English Corpus Linguistics: Vol. 9. Non-native Speech: A Corpus-based Analysis of Phonological and Phonetic Properties of L2 English and German*. Frankfurt, DE: Peter Lang.

Gut, U. (2012). The LeaP corpus: A multilingual corpus of spoken learner German and learner English. In T. Schmidt and K. Wörner (Eds.) *Hamburg studies on multilingualism: Vol. 14. Multilingual corpora and multilingual corpus analysis* (pp. 3-23). Amsterdam, NL: John Benjamins Publishing Company.

Gut, U. (2014a). The LeaP corpus: A phonetically annotated corpus of non-native speech. Retrieved from: https://sourceforge.net/projects/leapcorpus/files/?source=navbar

Gut, U. (2014b). *The LeaP corpus Manual.* Retrieved from: https://sourceforge.net/projects/leapcorpus/files/?source=navbar

Kilgour, J. (2017, April 10). *NITE XML toolkit homepages*. Retrieved from http://groups.inf.ed.ac.uk/nxt/index.shtml

Reppen, R. (2010). Building a corpus: What are the key considerations? In M. McCarthy & A. O'Keeffe. *The Routledge handbook of corpus linguistics* (pp. 59-65). New York, NY: Routledge.

Rohlfing, K., Loehr, D., Duncan, S., Brown, A., Franklin, A., Kimbara, I., Milde, J.-T., Parril, F., Rose, T., Schmidt, T., Sloetjes, H., Thies, A., & Wellinghoff, S. (2006). Comparison of multimodal annotation tools: Workshop report. *Discourse and Conversation Analysis*, 7. 99-123. Retrieved from: www.gespraechsforschung-ozs.de

Slavianova, E. (2007). *The LeaP corpus - Generating a relational database for linguistic query support* (Unpublished master’s thesis). Institute of Computer Science, University of Freiburg.

Gokturk, N. (2017). Mondly Learning Languages (Review). 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. 241-247). Ames, IA: Iowa State University.

SOFTWARE REVIEW

Mondly Learning Languages

Nazlinur Gokturk, Iowa State University

GENERAL DESCRIPTION

Mondly is a language learning software designed to teach core vocabulary and conversational skills needed to take part in everyday situations in 33 languages, including Spanish, French, German, Turkish, American English, British English, Japanese, Korean, Persian. Driven by the concept of 'gamification' in language learning, *Mondly* aims to provide its users with a chance to learn a foreign language in an enjoyable, practical, and effective way. Selected to be one of the top applications for language learning by Apple in 2015, it targets individual learners, travelers, and business professionals with a busy schedule. The software is available on website (https://www.mondlylanguages.com/), Apple's app store, and Google Play for the Android version. A free version grants access to first unit through membership. It offers its users several packs for a variety of deals such as a premium all languages lifetime access for $19.99, and one-month unlimited access to all languages for $9.99.

Mondly is comprised of twenty-three categories based on daily situations language learners are likely to encounter in real life such as going on a date, travelling, or public transportation. In each category, there are six lessons for teaching core words and phrases, along with the sentence formation exercises, one lesson for vocabulary practice, and one lesson for dialog practice. Below is a screenshot of a category with the lessons on a mobile device.

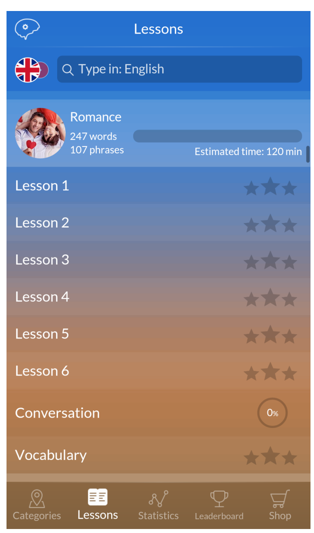


Figure 1. Eight lesson listed in a category

In addition to twenty-three categories, *Mondly* features a Conversational Chatbot (CC), which aims to enhance the learners' oral language skills through simulated dialogs with the computer. The CC includes three everyday situations such as ordering in a restaurant. Furthermore, an intelligent reporting system is incorporated into the software, which schedules the lessons to help learners to retain the skills they need without wasting time on review exercises and assigns weekly quizzes to the learners to measure their learning.

EVALUATION

Technological Evaluation

For the review, the Android version of Mondly was used. The course English for Turkish speakers was tested on a Samsung S5. The download of the app took approximately 2-3 minutes although it may vary depending on the user's internet connection speed. After signing up on the *Mondly* languages app, an individual premium license was purchased for $19.99 via of a debit card. Immediately after receiving the mail for the payment approval, all the categories became accessible on the app.

With respect to the interface, the intuitive design of the app is easy-to-navigate, which makes it appropriate for anyone with minimal computer knowledge (Figure 2). However, no manual is available on either website or app, although a couple of tutorials are presented for guidance on its YouTube Channel (https://www.youtube.com/channel/UCEAXQi1NM73KyCNNozI6jEA). There is also a "Feedback" icon at the "Statistics" tab on the app, which includes two tabs: a) report bugs, and b) ask for support. In the ask for support tab, the users are provided with step-by-step solutions for commonly encountered problems, and a comment box for their specific questions.

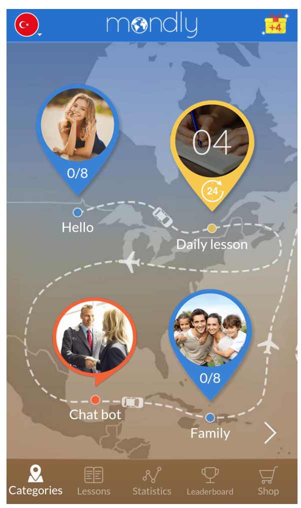


Figure 2. Home page

One of the most promising features of the software is its built-on automated speech recognition (ASR) technology in Conversational Chatbot. The ASR system in the CC not only recognizes voice input, but also gives replies with human voice in 33 languages. The system has been trained on natural spoken language data collected from multiple native speakers, and it automatically compares the learners' utterances to those of native speakers. However, it does not perform a detailed analysis of the learners' oral production in terms of phonological features, and thus it does not give any specific feedback regarding the learners' pronunciation.

Another feature of the software is its intelligent reporting system, which both keeps track of the learners' progress, and schedules the lessons based on the learners' quiz scores. Additionally, it helps the users to identify the areas they need to improve through detailed, yet easy-to-understand, graphs.

ACTIVITIES

The software consists of four different types of instructional activities: a) Conversational Chatbot, b) Dialog lessons, c) Flash card-based vocabulary exercises, and d) Translation-based exercises. Additionally, it offers a feature of Leaderboard where learners can compare their own performance in a lesson with learners all over the world. In this section, however, I will primarily describe the activities designed to improve the learners' oral language skills, which are Conversational Chatbot, and Dialog lessons.

In the Conversational Chatbot, the learners are presented with three different everyday situations such as ordering at a restaurant. After they choose the scenario, they are welcomed by an artificial conversation partner in the target language. To record their replies, the users are asked to tap and hold the microphone icon. Throughout the conversation, they are provided with a set of possible responses for each turn in the dialog, along with their pronunciation (see Figure 3). A help button is also available for the learners who want to check the translation of the replies in their native language. Although learners can craft their own responses, the system's ability to capture those statements is very restricted for now.

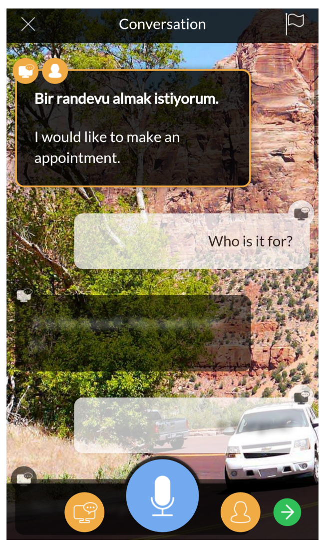


Figure 3. A dialog in Conversational Chatbot

As long as the learners respond to the questions accurately and clearly, they obtain scores for their replies, and the conversation proceeds. When they fail to produce accurate and clear utterances, they are presented with the question again and asked to repeat their response. Such a feature can improve both the learners' ability to produce meaningful utterances to sustain a conversation, and their pronunciation skills as they are supposed to pronounce their utterances accurately to maintain the conversation. At the end of each conversation, they are presented with a one-sentence learning outcome in their native language.

In Dialog lessons the users are presented with dialog scripts on a particular topic or situation such as holiday arrangements in a turn-by-turn fashion. In each turn, the learners are given the scripts both in their native and target language (Figure 4). In the activity, they first listen to a male and a female native speaker uttering the particular turn, and then they are asked to repeat the statement they have heard. In each conversation bubble, there appears two icons: one for the native speakers' recordings, and one for the learners' recordings (Figure 4). Those icons are also displayed at the bottom of the screen. In each turn, the learners can listen to both their own and native speaker's recordings for several times and compare their recordings to that of native speaker's. As long as they repeat the statements accurately, they are allowed to go further in the conversation, and can see the rest of the turns. At the end of each conversation they can play back both their own recordings and native speaker's recording, which would give them a chance to reflect on their own performance. As in Conversational Chatbot, the learners obtain scores for their performance and are provided with one-sentence learning outcome in the end of each dialog.

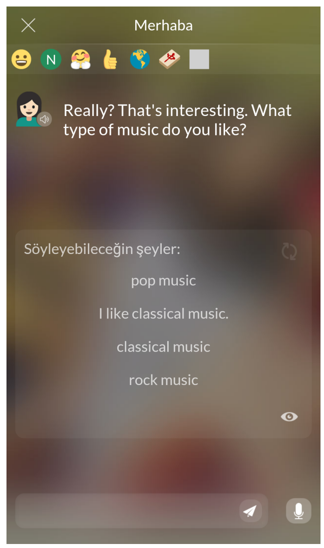


Figure 4. A dialog lesson in a course

TEACHER FIT (APPROACH)

Mondly has adopted Competency Based Language Teaching (CBLT) which focuses on "the design of work-related and survival-oriented language teaching programs for adults" (Richards, 2006, p. 40). It encompasses the teaching of the pre-determined language forms that would prepare the learners to communicate in real-life situations in the target language (L2). In line with CBLT, the focus of the software is on the achievement of learning outcomes, rather than the learning process (Richards, 2013).

Intended to provide the learners with opportunities to foster their speaking skills, Conversational Chatbot exemplifies the type of interaction realized between a person and a computer (Chapelle, 2003). Revisiting Ellis' (1999) work on interaction, Chapelle (2003) identifies three benefits of such an interaction type: a) enhanced input), b) attention to form, and c) negotiation of meaning. In this regard, CC appears to give opportunities for enhanced input through L1 translations, and for attention to form by requiring the learners to produce accurate and clear responses. Yet, what is missing in the CC is the opportunity for negotiating meaning. While learners can produce "comprehensible output" in CC (Swain, 1995), the nature of the output is restricted to a set of pre-fabricated responses. More creative opportunities would enable the learners to produce their own responses that would fit the context. As noted by Lin (2015) and Coniam (2008), exploring the ways of integrating a certain degree of negotiation into chatbots is certainly an area that merits further research.

In addition to the lack of opportunities for negotiating meaning, the native speaker speech in CC is slower than natural speech regardless of the language learning level (i.e., beginner, intermediate, and advanced). Although this may help low proficiency learners to comprehend the input (Krashen, 1985), the lack of authenticity in speech may lead to communication problems in real-life use (Cook, 1991; Ellis, 2008). A true adjustment of the speech pace based on the learner's proficiency would achieve this purpose. Another drawback of the CC is associated with the way it handles ungrammatical learner speech. As highlighted by Coniam (2008, 2014), chatbots targeted for second language learners need to be able to handle grammatical errors in learners' production to a certain extent. In this regard, CC appears to rely on phrases or expressions to sustain the conversation, and it ignores the grammatical accuracy of the learners' production. One way to mitigate this may be to incorporate a feature that would serve the purpose of recasts to CC. For example, upon receiving a statement like "I does like football" from the learner, the system may give a response like "Do you mean "I like football?". This would not only raise learners' awareness about language forms (Chapelle, 2003), but also give them a chance to produce accurate forms in their own speeches.

A useful feature of Dialog lessons is that it enables the learners to compare their own speech to that of native speaker's, and thus, to notice the gap between their own speeches and native speakers' speech. Schmidt (2001) suggests that 'SLA is largely driven by what learners pay attention to and notice in target language input and what they understand the significance of noticed input to be" (p. 4). One shortcoming of the Dialog lessons is that the software provides no explicit feedback on learners' pronunciation, except for asking for repetition. When the learner's response is not understood, the system asks for repetition until it is understood. As Levis (2007) points outs, feedback is an area where most computer assisted pronunciation teaching systems fall short as they are unable to perform an accurate analysis of learners' speech. Yet, given the importance of feedback in the enhancement of language learning (Chapelle & Jamieson, 2008; Gass, Behney, & Plonsky, 2013) the integration of a relatively more detailed explicit feedback feature to the system could improve the teaching of pronunciation in the software. Another drawback is associated with the lack of meaning focus. As highlighted by Chapelle (2001), a good language learning task should "direct the learners' attention to the meaning of the language" (p.56). This is certainly not taking place in such repetition-based dialog tasks. One solution would be to provide the learners with a set of responses, and to ask them to choose and repeat an appropriate response.

LEARNER FIT (APPROACH)

Considering the limited range of expressions and situations in Conversational Chatbot and Dialog lessons, *Mondly* may be appropriate for L2 learners at the initial stages of language learning. For those learners, the software can provide a safe environment to take part in meaningful conversations, and to practice their speaking skills in L2. However, for L2 learners with intermediate or higher levels of proficiency, the software would fall short of the expectations for two reasons. First, it offers only one dialogue practice on each situation in each course, and the Conversational Chatbot includes only three everyday situations. Additionally, according to company's website, *Mondly* adjusts the difficulty of the content based on the learners' language level. While it is true that a certain degree of adjustment is undertaken in the translation-based or flash-card activities, such an adjustment is not observable in the activities designed to improve the learners' speaking skills. For example, even in the advanced stage, the speech pace in both Chatbot and Dialog lessons is slow, and the dialog contents revolve around basic daily conversations such as making a hotel reservation. One way to address the intermediate or advanced L2 learners' needs would be to increase the variety of situations in Chatbot and Dialog lessons and to perform a true adjustment of the content depending on the proficiency level of the learner.

SUMMARY

Mondly language learning software would be best utilized by novice L2 learners, who would like to have a general introductory course in the target language possibly prior to a study-abroad, or business trip. The software features a Conversational Chatbot, and offers dialog lessons on everyday situations, which would be quite useful for L2 learners with limited command of target language. One way to enhance the effectiveness of the software for those learners would be to integrate a system of explicit feedback. On the other hand, for higher level L2 learners the affordances of the software for improving speaking skills would be probably insufficient in terms of both quantity, and quality. One alternative to address this shortcoming would be to extend the range of responses that the Conversational Chatbot can handle, and to increase both the quantity and variety of the conversations in dialog lessons so that the learners might have a chance to practice various conversations on a particular theme or situation. The software offers a premium license for US$19.99, which gives its users a life-time access to the courses in all 33 languages, and it is a good value for money for L2 learners who seek to have a general understanding in more than one language.

REFERENCES

Chapelle, C. A. (2001). *Computer applications in second language acquisition*. Cambridge: Cambridge University Press.

Chapelle, C. A. (2003). *English language learning and technology***.** Amsterdam: John Benjamins**.**

Chapelle, C. A., & Jamieson, J. (2008). *Tips for teaching with CALL: Practical approaches to computer-assisted language Learning*. New York: Pearson-Longman White Plains.

Coniam, D. (2008). Evaluating the language resources of chatbots for their potential in English as a second language. *ReCALL*, *20*(1), 98-116.

Coniam, D. (2014). The linguistic accuracy of chatbots: usability from an ESL perspective. *Text & Talk*, *34*(5), 545-567.

Cook, V. (1991). *Second language learning and language teaching*. Edward Arnold.

Ellis, R. (1999). *Learning a second language through interaction*. Amsterdam: John Benjamins Publishing Company.

Ellis, R. (2008). *The study of second language acquisition*. Oxford, UK: Oxford University Press.

Gass, S. M., Behney, J., & Plonsky, L. (2013). *Second language acquisition: An introductory course*. New York, NY: Routledge/Taylor and Francis Group.

Krashen, S. (1985). *The input hypothesis: Issues and implications*. London: Longman.

Levis, J. (2007). Computer technology in teaching and researching pronunciation. *Annual Review of Applied Linguistics*, 27, 184-202.

Lin, H. (2015). A meta-synthesis of empirical research on the effectiveness of computer-mediated communication (CMC) in SLA. *Language Learning & Technology*, *19*(2), 85-117.

Richards, J. C. (2006). *Communicative language teaching today*. Cambridge: Cambridge University Press.

Richards, J. C. (2013). Curriculum approaches in language teaching: Forward, central, and backward design. *RELC Journal*, *44*(1), 5-33.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge: Cambridge University Press.

Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics* (pp. 125–144). Oxford: Oxford University Press.

Jin, H. (2017). VoiceTube (Review). 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. 248-253). Ames, IA: Iowa State University.

SOFTWARE REVIEW

VoiceTube

Haeyun Jin, Iowa State University

INTRODUCTION

Previous literature on ESL pedagogy has shown that the traditional practice of teaching English pronunciation to ESL students tended to focus on isolated word- or sentence-level pronunciation (Hsieh, 2013; Pennington & Ellis, 2000) and restricted the teaching to isolated segments rather than the level of suprasegmentals (Lightbown & Spada, 2006). As opposed to such a decontextualized way of teaching pronunciation (Celce-Murcia et al., 2010), other research has emphasized the role of suprasegmentals in enhancing intelligibility (Avery & Ehrlich, 1992), and as areas considered more likely to affect communicative success (Derwing & Rossiter, 2003). Affected by such a redirection of focus, more emphasis was placed on the research and instruction on teaching intonation, stress, rhythm, and pause. Empirical research has noted that suprasegmental features, especially intonation should be taught at the discourse level (Chun, 2002; Jenkins, 2004; Levis & Pickering, 2004) and the traditional decontextualized way of sentence-level teaching may not be effective (Levis, 1999).

Shadowing for suprasegmentals

Shadowing is often suggested as one effective teaching and learning method to promote the learning of suprasegmental features within the discourse context. To follow the definition by Luo et al. (2010), the basic skill of shadowing is to immediately follow the utterance produced by the native speaker "as closely as possible." Since learners should remember the speech input and reproduce the original intonation, speed, and stress pattern exactly the same way as the original speaker in real time, it is counted as a highly cognitive action rather than a mere automatic parroting (Hamazah & Miko, 2010). Although this method was originally used for training beginner simultaneous interpreters (Lambert, 1992), it currently has become a widely employed pedagogical technique for improving English pronunciation (Hamazah & Miko, 2010). Furthermore, with regard to shadowing strategy and the teaching of suprasegmentals, Celce-Murcia et al. (2010) hold that shadowing can be considered an effective teaching method for imitating native speech intonation patterns at the discourse level. In this respect, *VoiceTube* can function as one tool for shadowing and ultimately for improving English suprasegmentals in discourse context.

SOFTWARE DESCRIPTION

VoiceTube is a free web-based software that allows practice learning English pronunciation via video resources. As the largest English learning community in Taiwan, it provides over 40,000 videos with 17 different channels of video sources including TED talks, BBC Learning English, CNN Student News, movie clips, music videos, and others with 18 broad categories of topics. One convenient feature of this software is that it has iOS and Android mobile applications. Thus, it is a highly compatible software with current practice of learning English where accessibility and convenience is valued. One helpful feature that distinguishes *VoiceTube* from other seemingly similar resources for authentic English speech such as *TedTalk* or *Podcast* is that this software provides the shadowing tool in its speaking section along with its video section. Thus, it can be counted as a comprehensive system for practicing shadowing technique through listening, speaking and vocabulary.

HOW IT WORKS

Learners can start by choosing their favorite videos from the list. As seen in Figure 1, they can choose the video by channels of interest (17 channels), their levels (basic/intermediate/advanced), and topics (18 categories). Also, they can choose videos by accent (US/UK/AU/Other) and duration.

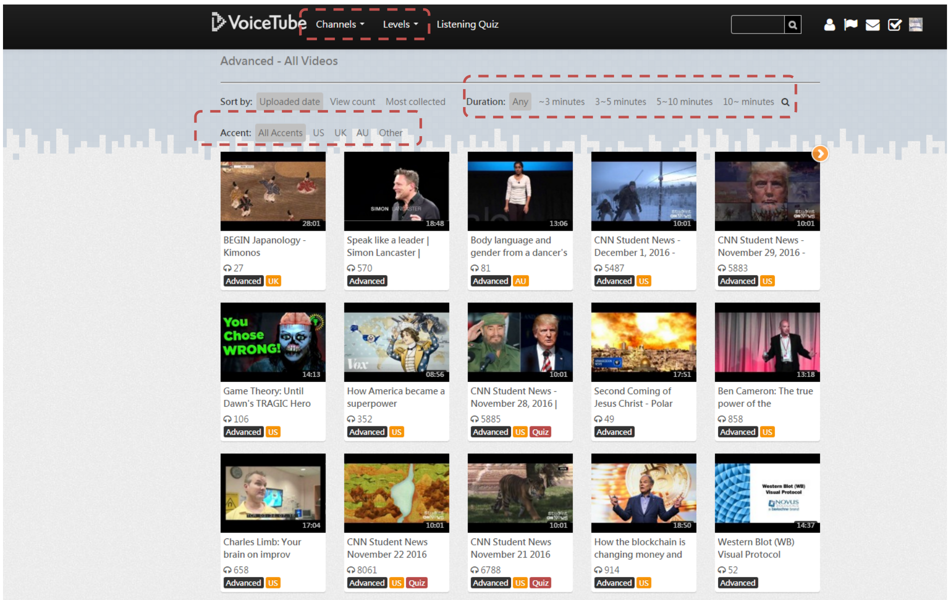


Figure 1. VoiceTube Webpage

The aim of this step is to find the video with the specific accent that each learner prefers to practice. *VoiceTube* provides "other" and "AU” (Australian English) categories in addition to General American English and Received Pronunciation (Levis, 2005). The "Other" category includes the videos of various English varieties such as Philippine English and even includes nonnative speakers of English such as Mexican, Chinese, and Swedish. It should also be noted that learners should select the video which is not too difficult for their levels. Since the ultimate goal of shadowing is to imitate the speech and its suprasegmental features, lack of comprehension should not interrupt the speaking process (Hamada, 2014; Hamazah & Miko, 2010). If learners are in lower levels, it is recommended that they choose videos with shorter durations.

The next step is the listening stage in which learners watch the video in the form of connected speech. As mentioned above, in order to prevent the situation where lack of comprehension skills impede speaking, lower level learners can choose to watch with the help of subtitles provided in English and Chinese and adjust the speed of the speech. As can be seen from Figure 2, they can also refer to the sentence by sentence caption provided right next to the video player. Learners can watch the full video several times or choose to repeat each sentence by clicking the green colored captions.

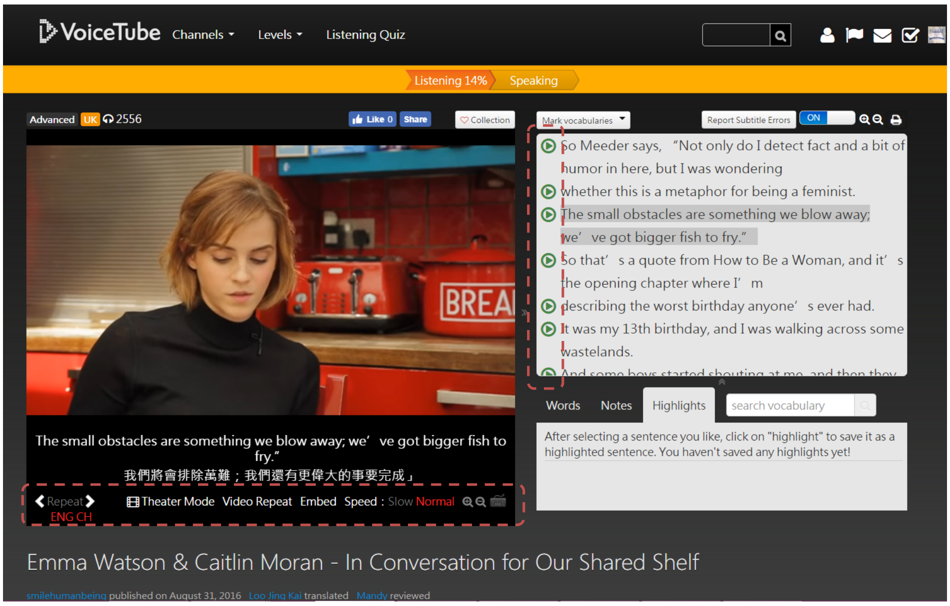


Figure 2. The Listening Stage

The goal of this stage is to listen to the video several times so that learners can become familiar with the speed and the intonation pattern of the speech.

The last step is the speaking stage, or the actual shadowing stage. By clicking on the captions (Figure 3), learners can play the original speech sentence by sentence. Then, they can start imitating the model speech and record it by clicking on the microphone. This process occurs in real time while you are listening to the model speech. If learners stop recording, the player automatically gives them the recording of their speech followed by the original model speech in the video. If further comparison is needed, they can click on the button right to the microphone ("*compare my recording to the video*") and listen to the pair of their own recording and the model speech as much as they want.

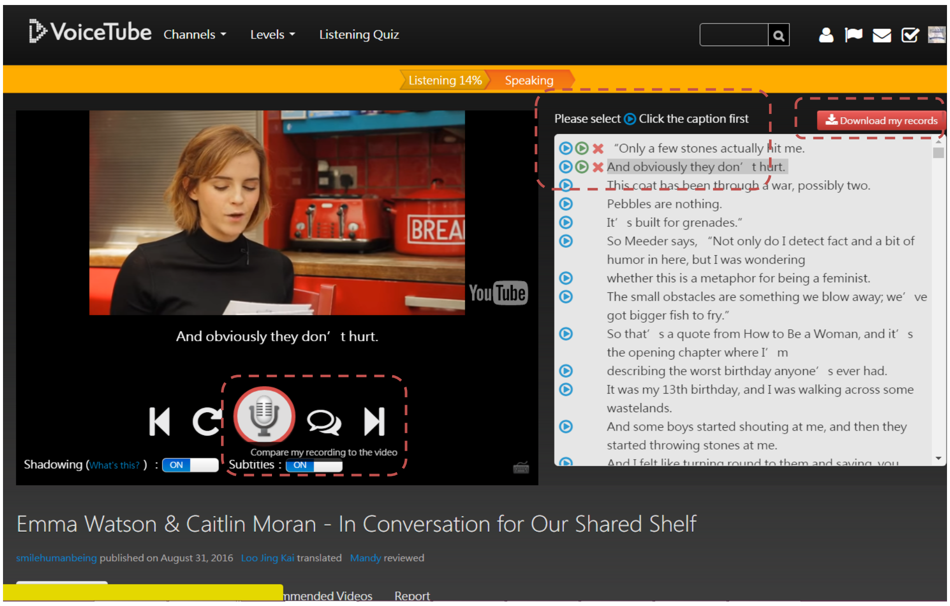


Figure 3. The Shadowing Stage

The aim of this stage is to follow the model speech imitating its intonation, stress patterns, pause, and speed. By comparing their own speech to the model sentence by sentence, learners can effectively figure out their weaknesses in terms of suprasegmental features and promote learning. They can also download their own speech in mp3 format or keep track of their practice in the account management page.

STRENGTHS

Authentic, discourse-level input

VoiceTube provides a wide array of authentic, communicative, and highly contextualized speech input. The videos can be counted as authentic in that they are the most up-to-date, real life speech about a wide variety of topics. Moreover, the authenticity of this software also comes from the fact that it includes nonnative English speech models along with the native varieties, which is the reality that ESL learners should face in the real life. Considering that current materials for teaching pronunciation are mostly based on *nativeness principle* depending exclusively on American English or Received Pronunciation (RP) (Levis, 2005), those nonnative speech models provided by *VoiceTube* are somewhat inspiring for second language pronunciation teaching. Also, *VoiceTube* is a good source of learning English suprasegmentals since it provides models for discourse-level connected speech. From the previous literature, it is widely accepted that suprasegmental features should be taught at the discourse level rather than in relatively decontextualized and isolated setting such as phrase- or sentence-level teaching (Chun, 2002; Jenkins, 2004; Levis & Pickering, 2004). Thus, all in all, *VoiceTube* is a purposeful tool for learning English pronunciation, especially intonation, because it provides authentic, discourse level input for practice.

Effective functionalities for shadowing

The second advantage of *VoiceTube* comes from its highly effective functionalities as a tool for shadowing. First, it provides a comprehensive system for shadowing composed of listening, speaking, and vocabulary. The specific feature that distinguishes this software from other sources is that it also supports learner production. By way of the recording tool built in the video player, learners can actually record their speech and readily compare it with the model speech. Considering that seemingly similar software such as *TedTalk* only works for listening, *VoiceTube* provides a more effective platform to practice shadowing. Moreover, it provides other convenient and purposeful utilities for promoting shadowing. It is claimed from the previous literature that shadowing is effective after learning target contents, and that learners can use this practice with a lower cognitive burden. (Hamada, 2014; Hamazah & Miko, 2010). Therefore, comprehension of the content should be preceded for successful shadowing of the speech. *VoiceTube* has several functions to enhance learner's comprehension of content. For instance, learners can choose to watch the video with subtitles (either target language or their L1) and captions. They can also adjust the speed of the speech and access the meaning for vocabulary. In this respect, these functionalities can ultimately help prevent the situation where lower level learners' lack of comprehension impedes the shadowing of the speech.

Highly individualized setting

The last strength of this software in terms of the learner fit (Chapelle, 2001) is that it provides a highly individualized setting for learning English pronunciation by giving a wide range of choices for videos by learning purposes, levels, topics, and even target English varieties. It enables the learners to purposefully practice according to their own specific aims. This software is also helpful in that learners can track their learning process, save the videos, save their recordings, and create their own wordbank and notes.

POSSIBILE IMPROVEMENTS

Automatic Speech Recognition (ASR) and feedback

One limitation of *VoiceTube* is that it does not provide acoustic analysis of the learner's recorded speech, or feedback. Current one-way production of recording can be less motivating since it is not interactive. Lower level students, in particular, might need help understanding acoustic features of their own speech compared to the model speech. However, current system does not support speech recognition nor give analysis for the learners. In this respect, ASR can help ESL pronunciation by providing one-to-one feedback (Chiu et al., 2007; Jung, 2011). Therefore, adding speech recognition technology to the current system will help rendering the practice more interactive. If the system offers interaction which resembles the real-life communication more, it will be more motivating for the learners (Alsatuey, 2011).

Expanded levels of shadowing

The second suggestion relates to the way shadowing functions in the software. Currently, it provides a sentence-by-sentence shadowing option only. However, if learners can gradually expand their level of shadowing (e.g. one sentence → multiple sentences → paragraph), it will be more purposeful and meaningful for practicing suprasegmental features in discourse level. In this respect, Levis & Pickering (2004) noted that intonation in discourse is consistently and systematically different from intonation patterns in isolated phrases and sentences. Thus, it would make more sense if the format of shadowing matches the format of the actual speech, which is the connected speech.

Other technical limitations

The last limitation relates to other technical shortcomings of this program. First, the current version only supports English and Chinese subtitles since it was originally developed in Taiwan. Considering that lower level students might need L1 to understand the content of the video, more L1 varieties should be added for wider usability of the software. Also, despite the provision of nonnative model speech in the current system, which is distinct from other softwares, it still has limitations in that they provide relatively restricted varieties of videos for AU and Other category. Thus, I would suggest that they add more sources for nonnative model speech.

REFERENCES

Alsatuey, M. C. (2011). Perceived benefits and drawbacks of synchronous voice-based computer-mediated communication in the foreign language classroom. *Computer-Assisted Language Learning*, *24*(5), 419–432.

Avery, P., Mackin, R., & Ehrlich, S. L. (1992). *Teaching American English pronunciation*. Cary, NC: Oxford University Press.

Chapelle, C. (2001). *Computer applications in second language acquisition: Foundations for teaching, testing, and research.* Cambridge: Cambridge University Press.

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

Chiu, T.-L., Liou, H.-C. & Yeh, Y. (2007). A study of web-based oral activities enhanced by automatic speech recognition for EFL college learning. *Computer-Assisted Language Learning*, *20*(3), 209–233.

Chun, D. (2002). *Discourse intonation in L2: From theory and research to practice*. John Benjamins, Philadelphia.

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

Lambert, S. (1992). Shadowing. *Meta, 37*(2), 263-272.

Levis, J. M. (1999). Intonation in theory and practice, revisited. *TESOL Quarterly*, *33*(1), 37-63.

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

Levis, J. M., & Pickering, L. (2004). Teaching intonation in discourse using speech Visualization technology. *System*, *32*(4), 505-524.

Lightbown, P. M., & Spada, N. (2006). *How languages are learned*. Oxford: Oxford University Press.

Luo, D., Qiao, Y., Minematsu, N., Yamauchi, Y., & Hirose, K. (2009). Analysis and utilization of MLLR speaker adaptation technique for learners' pronunciation evaluation. In *INTERSPEECH* (pp. 608-611). Brighton, UK.

Hamada, Y. (2014). The effectiveness of pre- and post-shadowing in improving listening comprehension skills. *The Language Teacher, 38*(1), 3-10.

Hamzah Md Omar, M., & Miko, U. (2010). Using 'A shadowing' technique to improve English pronunciation deficient adult Japanese learners: An action research on expatriate Japanese adult learners. *Journal of Asia TEFL*, *7*(2), 199-230

Hsieh, K. T., Dong, D. H., & Wang, L. Y. (2013). A preliminary study of applying shadowing technique to English intonation instruction. *Taiwan Journal of Linguistics*, *11*(2), 43-65.

Jenkins, J. (2004). Research in teaching pronunciation and intonation. *Annual Review of Applied Linguistics*, *24*, 109-125.

Jung, H.-K. (2011). The correction of learner's English pronunciation errors through speech recognition reading program. *Multimedia-Assisted Language Learning*, *14*(3), 291–314.

Pennington, M. C., & Ellis, N. C. (2000). Cantonese speakers' memory for English sentences with prosodic cues. *The Modern Language Journal*, *84*(3), 372-389.

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.

Karatay, Y. (2017). YouGlish.com (Review). 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. 254-259). Ames, IA: Iowa State University.

WEBSITE REVIEW

YouGlish.com

Yasin Karatay, Iowa State University

INTRODUCTION

There has been increasing interest in pronunciation teaching and learning in parallel with recent developments in technology. Thanks to these developments, we have witnessed a number of applications and websites focusing on pronunciation. As Levis (2007) stated, computers can be utilized for individualized instruction, frequent practice through listening discrimination and focused repetition exercises, and automatic visual support. In this respect, *YouGlish,* a website which relies on YouTube videos to present target vocabulary through videos from a variety of genres, offers an invaluable resource for English language learners (ELLs) to improve their pronunciation.

OVERVIEW OF THE WEBSITE

This website was first deployed as *YouPronounce* in February, 2015, then renamed as *Youglish* in January, 2016 by Dan Barhen, a software engineer from Paris. The basic idea behind this website is to provide users with fast and unbiased answers about how English is spoken by real people and in context instead of what's prescriptively correct. According to the updated information on the website, *YouGlish* currently supports more than 300,000 terms which can be searched in three different accents of English (British, American or Australian). Although it only provides results in English, *YouGlish* offers five language options (English, Spanish, German, Portuguese, Italian, and French) for explanations.

How It Works

YouGlish illustrate for users how to pronounce English words by presenting them in natural speech from native or fluent speakers by making use of over ten million YouTube videos. In other words, users can be exposed to target words that are represented in natural contexts. In our email correspondence, Dan Barhen stated that listening to real people pronouncing an utterance in a real context is much more effective than a regular dictionary approach. He reported that technically, *YouGlish* uses an algorithm that processes YouTube videos by taking into account numerous parameters such as the caption/audio synchronization, video/audio quality, video statistics (number of views, likes, etc.), video quality, restriction (regional, on site), language, accent, user inputs and much more to figure out how each video should be indexed and ranked.

Basically, *YouGlish* allows users to type any word, phrase, or sentence in the search tab to watch a video where someone is pronouncing that keyword. After the users get the results, the video automatically starts from the relevant point, which is a great feature that helps you find the right place. If the users want more examples of the target word or phrase, they simply click on an arrow button to go the next video result (see Figure 1).

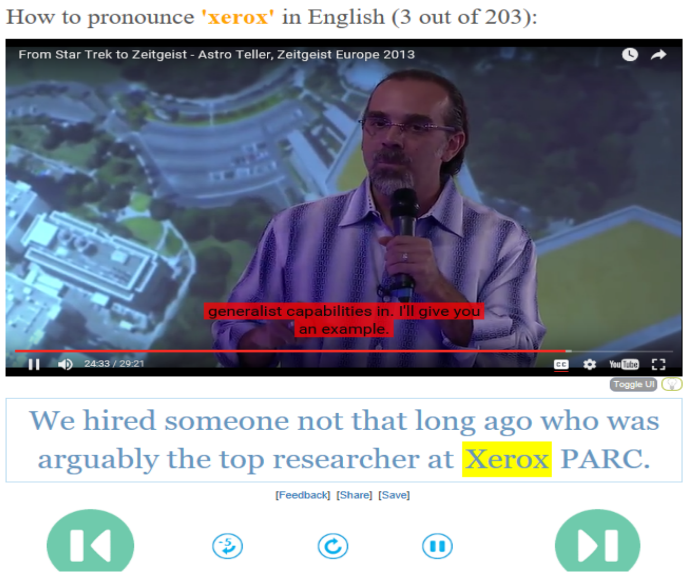


Figure 1. Highlighted search terms on YouGlish.

As seen in Figure 1, the site presents the video with the target word color-coded in the transcript. Another important function is that all the words in every transcript are hyper-linked to different videos making it relatively easy for users to navigate between target words and videos.

Search Tab

YouGlish allows users to search any kind of utterances to check their pronunciation. They can search a word (i.e., Xerox), a phrase (i.e., a great deal of), a sentence (I'd appreciate it.). It can even search parts of sentence such as 'the tools but', or 'is essentially' if the users want to listen to how two different words are pronounced in different contexts. Additionally, users can use hashtags to look for a word in a certain context, e.g., iowa#education, iowa#TedTalks, button#computer (see Figure 2). In this case, the target word is the one before the hashtag, and the video context to listen to the target word in is the one after the hashtag.

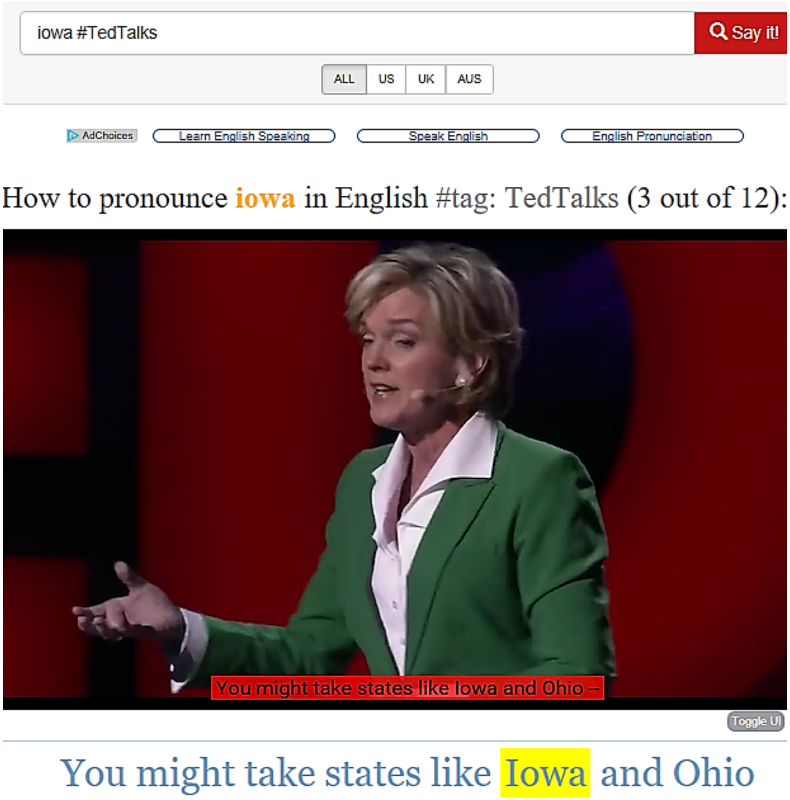


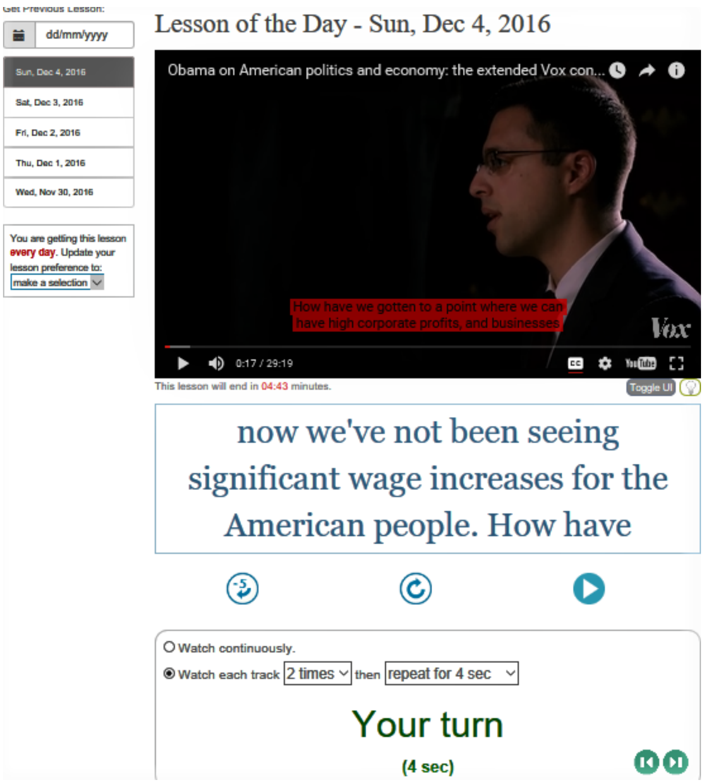
Figure 2. Using hashtages to target particular content.

For each result, users can make use of three sections, in addition to the videos, dedicated to *Nearby words* where students can click on the hyperlinked words similar to the target word, *Phonetic* where users can see the Phonetic transcription of the target word, and *Tips to improve English pronunciation* where users are provided with several suggestions regarding pronunciation.

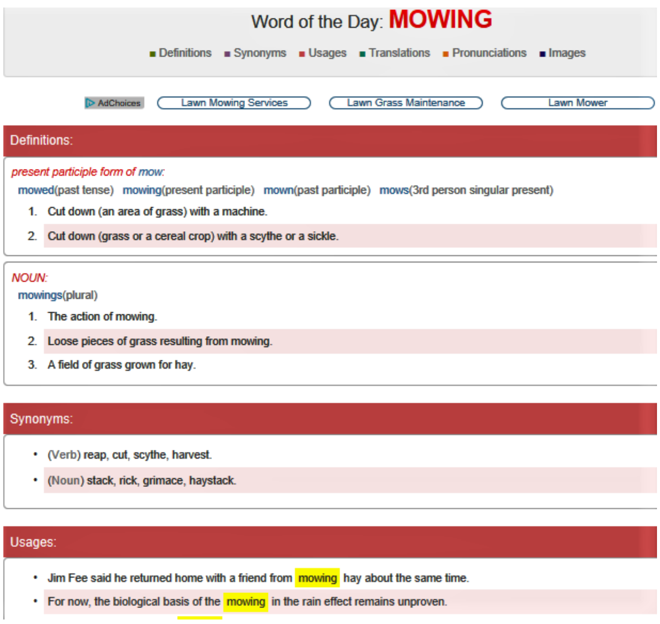
YouGlish can also be run in a *'Restricted Mode'* that will block any inappropriate content from being displayed. If teachers plan to use *YouGlish* in their classes, they can activate this feature by just clicking on the *'Restricted Mode'* button at the bottom of any page.

Referring to what Levis (2007) suggested as the ways that computers can be used in pronunciation, there are three other functions this website offers its users.

Lesson of the day: After signing up, based on the settings, users can receive emails of a randomly-selected video everyday, twice a week, once a week, or once a month. What makes this option different from what has been described so far is the different settings offered in each lesson (Figure 2). Learners can watch the video continuously or adjust the duration and repetition times of the utterance to listen to a specific section of the video. For example, they can watch every four seconds of the video twice and then they will have a four second interval to repeat what they have heard.



Word of the day: Similar to the '*Lesson of the day'*, users can also receive emails regarding a randomly-selected word every day, twice a week, once a week, or once a month. Users can make use of the definitions, synonyms, usages (10 different sentences), translation to 52 different languages, pictures/images if available, and pronunciation (videos) of the target word. In the timeline tab, since the website tracks the user's content, users can also refer back to the previous words of the day (see on the right).



My content: Under every video, users are provided with a 'save' button for future reference. If saved, the users can watch these videos from 'My content' section whenever they want.

EVALUATION

This free, user-friendly website is rich in content and has great potential for aiding pronunciation improvement. Considering that the users have the opportunity to focus on problematic aspects of their own pronunciation by moving at their own pace, *YouGlish* can also promote autonomous learning (Lu, 2010; Hanna & Gao, 2016). In addition to self-paced learning, it should also be noted that it offers a stress-free environment (Roed, 2003). Although there are a variety of software applications available to teach pronunciation, *YouGlish* stands out among them because it presents target vocabulary in authentic contexts. It provides input not only in segmental level but also at the suprasegmental level such as sentence stress, intonation, rhythm and word linking because learners listen to speech in natural context uttered by native speakers, which is of great importance in pronunciation (Derwing & Rossiter, 2002).

However, there are some issues which need improvement. First of all, YouGlish does not give its users the opportunity to produce the speech they hear, record it, and receive feedback. As Kissling (2013) underscores, input, practice, and feedback are the factors that affect a learner's pronunciation improvement most. However, *YouGlish* is one of the best websites by far regarding the input provided and the potential for practice, even though it does not provide feedback. As Flege and Wang (1989) argued, learners should be helped noticing how they are doing by being given immediate feedback.

Another limitation is that *YouGlish* does not provide phonetic transcription of the words on the 'word of the day' page although it provides this when the learners watch the video separately from the homepage. Also, since it relies on only YouTube, the videos are not viewable in some countries where YouTube is blocked.

Considering that it is a relatively recent website, I should say that it has the potential of being one of the leading websites in pronunciation learning. As a nonnative speaker of English, I believe many learners would be extremely happy when they find what this website can do for them.

ABOUT THE AUTHOR

Yasin Karatay holds a BA in English Language Teaching from Anadolu University and MA in TEFL from Bilkent University in Turkey. He has taught a variety of English courses at the tertiary level, including academic reading and writing, grammar, and technical English. He has presented at local and international conferences, including IATEFL and EuroCALL. He has worked as an EFL instructor and assistant director at Duzce University in Turkey. He is currently a PhD student in the Applied Linguistics and Technology program and an ESL instructor at Iowa State University. His research interests are automated writing evaluation, CALL use in material development and assessment, and second language acquisition.

REFERENCES

Derwing, T. M., & Rossiter, M. J. (2002). The effects of pronunciation instruction on the accuracy, fluency, and complexity of L2 accented speech. *Applied Language Learning, 13*(1), 1-17.

Flege, J. E., & Wang, C. (1989). Native-language phonotactic constraints affect how well Chinese subjects perceive the word-final English/t/-/d/contrast. *Journal of Phonetics, 17*(4), 299-315.

Hanna, B. E., & Gao, Y. (2016). Exploring optimal pronunciation teaching: Integrating instructional software into intermediate-level EFL classes in China. *CALICO Journal, 33*(2), 201-230. doi: http://dx.doi.org.proxy.lib.iastate.edu/10.1558/cj.v33i2.26054

Kissling, E. M. (2013). Teaching pronunciation: Is explicit phonetics instruction beneficial for FL learners? *The Modern Language Journal, 97*(3), 720–744.

Levis, J. M. (2007). Computer technology in teaching and researching pronunciation. *Annual Review of Applied Linguistics, 27*, 184-202. Retrieved from http://search.proquest.com.proxy.lib.iastate.edu/docview/85669655?accountid=10906

Lu, D. (2010). A salutary lesson from a computer-based self-access language learning project. *Computer Assisted Language Learning, 23*(4), 343-359. Retrieved from http://search.proquest.com.proxy.lib.iastate.edu/docview/772290692?accountid=10906

Roed, J. (2003). Language learner behaviour in a virtual environment. *Computer Assisted Language Learning, 16*(2–3), 155–172.

Lockwood, J. (2017) NORM the vowel normalization and plotting suite (Review). 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. 260-263). Ames, IA: Iowa State University.

WEBSITE REVIEW

NORM: The Vowel Normalization and Plotting Suite

Jeremy Lockwood, Iowa State University

INTRODUCTION

In sociophonetic research investigating language variation, particularly with regard to the perception and production of vowels, vowel normalization methods have been developed to be able "to compare vowel realizations by different speakers in meaningful linguistic and sociolinguistic ways" (Thomas & Kendall, 2007, p. About Vowel Normalization). Ideally, the vowel normalization methods preserve phonemic and sociolinguistic variation while minimizing the physiological variation between speakers (Adank, 2003). Thomas (2010) listed four goals for normalizing formant data: "to eliminate variation caused by physiological differences among speakers (i.e., differences in vocal tract lengths); preserving sociolinguistic/dialectal/cross-linguistic differences in vowel quality; preserving phonological distinctions among vowels; [and] modelling the cognitive processes that allow human listeners to normalize vowels uttered by different speakers" (p. 161). Several vowel norming methods have been developed, and the tool being reviewed here allows one to implement nine normalization methods on formant data. Each normalization method is described in the application along with their advantages and disadvantages as well as the reason for using them.

The website was created to facilitate manipulation, normalization, and plotting of vowel formant data for sociophoneticians, phoneticians, and sociolinguists (Thomas & Kendall, 2007). The design of the website allows for the quick application of the methods for normalization so that the results from each manipulation can be compared and visualized. The website emphasizes knowing the purpose of one's research as well as the goals for vowel formant normalization. All of these aspects are discussed over the several tabs on the website, and the creators provide extensive detail for each part of the tool to help researchers be better informed as well as to make the tool, which is a bit inaccessible, easier to understand and use.

CREATORS OF THE WEBSITE

NORM: The Vowel Normalization and Plotting Suite was created by Tyler Kendall and Erik Thomas. Tyler Kendall is an associate professor at the University of Oregon. Kendall's research has been in sociolinguistics with interests in social and cognitive aspects of language variation and change. He tends toward quantitative research in sociolinguistics, although he also does research incorporating computational linguistics, corpus linguistics, lab phonetics, and psycholinguistics. Kendall also developed the Sociolinguistic Archive and Analysis Project (SLAAP) and the Online Speech/Corpora Archive and Analysis Resource (OSCAAR), which are web-based language archives. Erik Thomas is a professor at North Carolina State University whose interests involve sociophonetics focusing on acoustic analysis of language variation, cognition of language variation, and variation in English dialects. His current work is in new dialect formation.

WEBSITE CONTENT AND LAYOUT

The website has a fairly basic layout. A screenshot of the homepage can be seen in figure 1 below. There are six tabs that one can click on to navigate the site: About NORM, About Vowel Normalization, Normalization Methods, How to Use NORM, Bibliography, and NORMalize!.

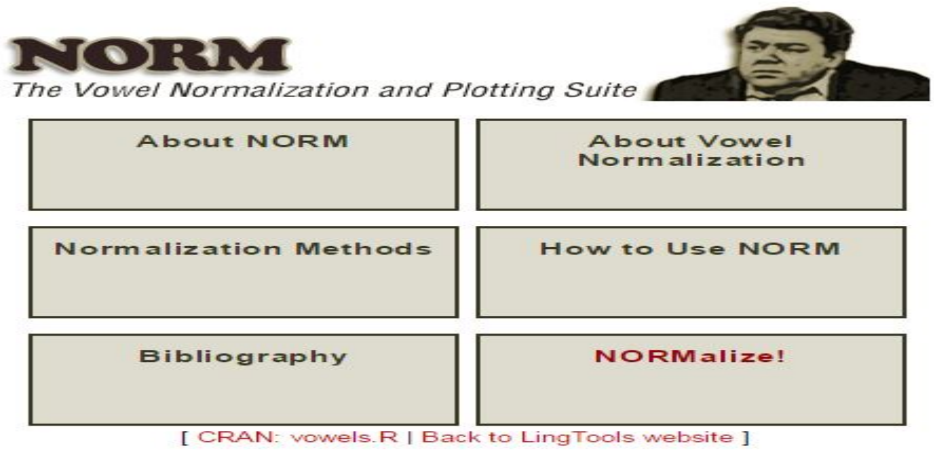


Figure 1. NORM homepage found at URL http://lingtools.uoregon.edu/norm/index.php.

The About NORM tab contains information describing the tool and what it is intended for. The headings under this tab are *About NORM*, *How NORM Works*, *About NORM's Scaling*, and *How to Cite NORM*. In each section there are hyperlinks to direct the user to other tabs within NORM and other websites. The About Vowel Normalization tab contains information under only the title of the tab. There, the creators provide information on vowel normalization, goals for normalization, how differing research may focus on certain goals over others, and two images showing the effect of one of the normalization methods on the acoustic vowel charts of two speakers. Under this tab, the creators provided citations of research that used and reviewed the normalization techniques. The Normalization Methods tab provides information on each of the normalization methods that the website allows users to apply to their data. Above it was mentioned that the website is able to carry out nine normalization methods; however, these nine methods were derived from five main methods: Bark Difference Metric, Labov, Lobanov, Nearey, and Watt & Fabricius. For each method, the creators provide information about the method and what it does; how the method affects formant data; advantages and disadvantages of the method; an image of the output data; and if there are multiple versions of a method, there is additional information on each. The How to Use NORM tab has two main headings: *Using NORM* and *Preparing Your Data*. The first heading forecasts that to use the NORM application, the data need to be in a certain format. The second heading, *Preparing Your Data*, describes the options one has when considering how to set up their data, and it provides two templates as well as several sample sets of data that can be used to test the application. The bibliography tab shows the sources that were used to inform the content of the other pages of the website. The last tab on the NORM homepage is the NORMalize! tab. The screenshot below in figure 2 shows the layout of the part of the website that serves as the tool for normalizing formant data.

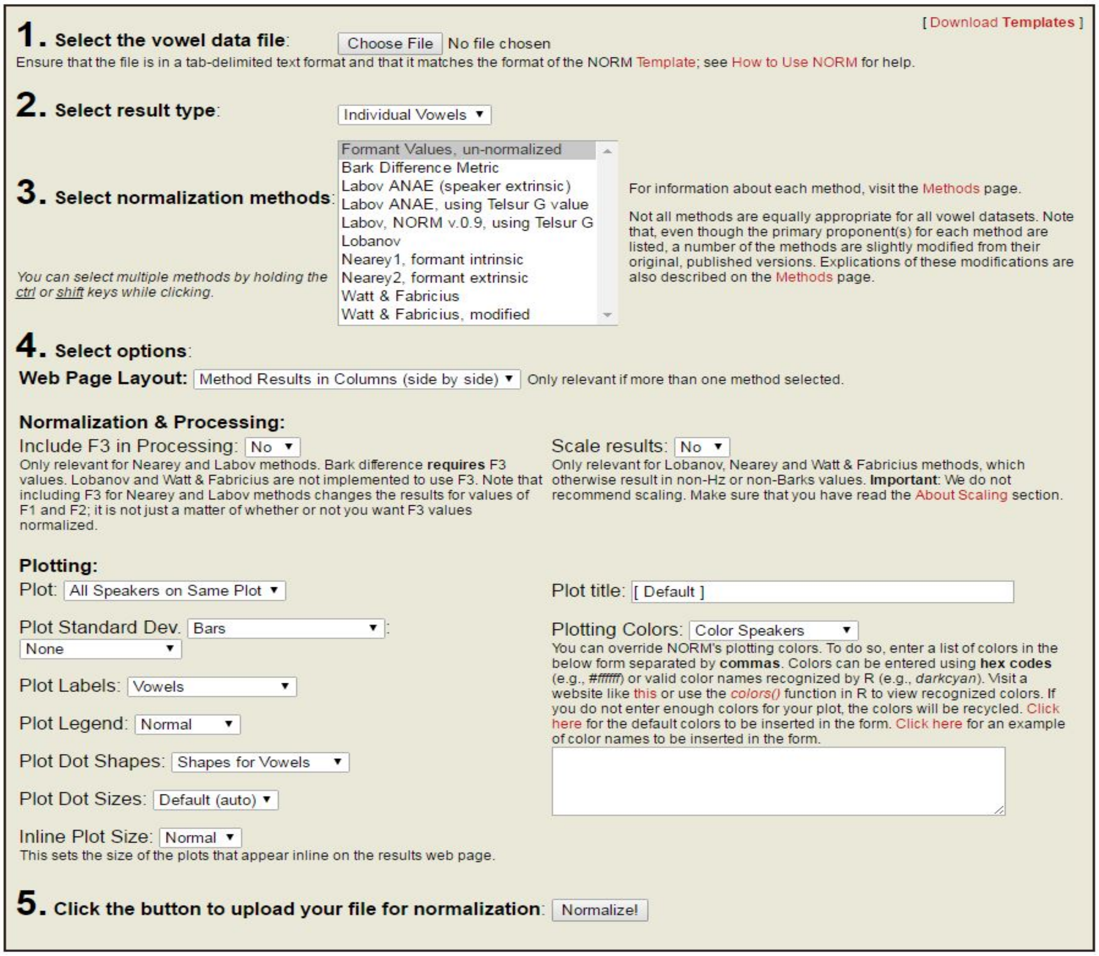


Figure 2. *NORMalize! Page Found at URL* http://lingtools.uoregon.edu/norm/norm1.php.

The NORMalize! page above shows the options and steps available to normalize or manipulate. Step one allows the user to choose the files to use with the application. The section reminds the user of the formatting requirements for using the tool, and it provides hyperlinks to other sections of the website where the user can find the templates for setting up the data and how to use the tool. The second step in the process asks the user to select the type of results the user wants, whether the user seeks to analyze individual vowels, speaker means, or group means. The third step gives the user the option to select the normalization method for the data. It provides links to the page that provides information on the different types of normalization methods, and it allows the user to select more than one normalization method to apply to the data. Step four allows the user to select the additional options for the output of the website, starting with the layout of the webpage, inclusion of the third formant and information to guide that decision, scaling of results, and different plotting options. Step five serves to carry out the selected normalization method, and it completes the process.

EVALUATION

In evaluating this website, three criteria will be discussed based on an evaluation of website content and structure—content, interaction, and navigation (Bauer & Scharl, 2000). The content of the website is very specialized, and the tool is clearly marked as appropriate for sociophoneticians, phoneticians, and sociolinguists. One without prior knowledge about normalization of formant data or why one might seek to compare vowel formant data would likely become overwhelmed quickly by the amount of information about each method and the discussion about advanced manipulation through creating R scripts. The advanced programming knowledge is not required for using this tool, but there are other requirements for using the tool such as knowledge about acoustic analysis and acquiring formant data. The tool assumes much of that knowledge in its users. The tool provides detailed information about each normalization method and the advantages and disadvantages of each as well as sources for additional reading. Without those resources, this tool would be very difficult to use meaningfully. The interaction between the user and the tool involves the user being conscious of the website's abilities. The user has the opportunity to download templates and sample data sets to test with the tool. The user must upload data in the form of a tab-delimited text document, then select from the several options that the tool provides. The website also allows the user to select the layout of the output as well as labelling and color coding options. The navigation aspect could be improved, although with some experience with the website, the user can learn how to navigate the tool more easily. The first area for improvement could be in the difference between the tabs on the tool's homepage and the tabs on the rest of the pages. The homepage contained six main buttons; however, the top of each page after the homepage has seven buttons, and the labels vary from the labels on the homepage. This seems to be only a minor inconvenience. As a new user, navigation was difficult primarily because the tool would suggest that, to find additional/clarifying information regarding more advanced topics, to click on link that would show another page on the website creating a bit of a rabbit hole feel to the website.

Considering the audience for this tool, which was stated as sociophoneticians, phoneticians, and sociolinguists, it seems that the tool facilitates quick and meaningful processing of vowel formant data. The tool seems to be valuable, although user training would likely be needed and very beneficial especially to less experienced researchers.

REFERENCES

Adank, P. (2003). *Vowel Normalization: A Perceptual-Acoustic Study of Dutch Vowels*. PhD thesis, University of Nijmegen.

Bauer, C. & Scharl, A. (2000). Quantitative evaluation of web site content and structure. *Internet Research*, *10*(1), 31-44.

Thomas, E. (2010). *Sociophonetics: An Introduction*. Basingstoke, U.K./New York: Palgrave Macmillan.

Thomas, E. & Kendall, T. (2007). NORM: The vowel normalization and plotting suite. [Online Resource: http://ncslaap.lib.ncsu.edu/tools/norm/]

Phng, S.W. (2017). *American English pronunciation tutor* (Review). 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. 264-269). Ames, IA: Iowa State University.

SMARTPHONE APP REVIEW

The *American English Pronunciation Tutor*

Sock Wun Phng, Iowa State University

INTRODUCTION

American English Pronunciation Tutor (Language Arts Digital, 2016) is a smartphone application (app) developed by Language Arts Press in 2016 (Language Arts Press Products). It is available for devices running the Android or iOS operating system and can be downloaded for free from Google Play or iTunes (Language Arts Press Products). Language Arts Press claims that by using the app, users can (1) develop pronunciation of vowel and consonant sounds, accuracy and awareness of grammatical endings, word-level stress, sentence-level stress, and rhythm, (2) improve fluency and grammatical awareness, and (3) develop clear and confident speech. The target population stated for the app is high beginner to advanced learners of American English (Language Arts Press Products).

OVERVIEW

Ten units are included in the app (Figure 1). Units 1 to 8 cover segmentals, while Units 9 and 10 cover suprasegmentals. In the free version of the app, only Unit 1 is available to the user. In order to access the other nine units, the user must upgrade to the Pro version of the app for $5.99. Fortunately, Unit 1 has the most lessons out of all of the units. It starts with an introduction to front vowels, followed by five lessons on five front vowels (Figure 2). The other units follow the same basic structure: there is an introduction to the unit and a few lessons.

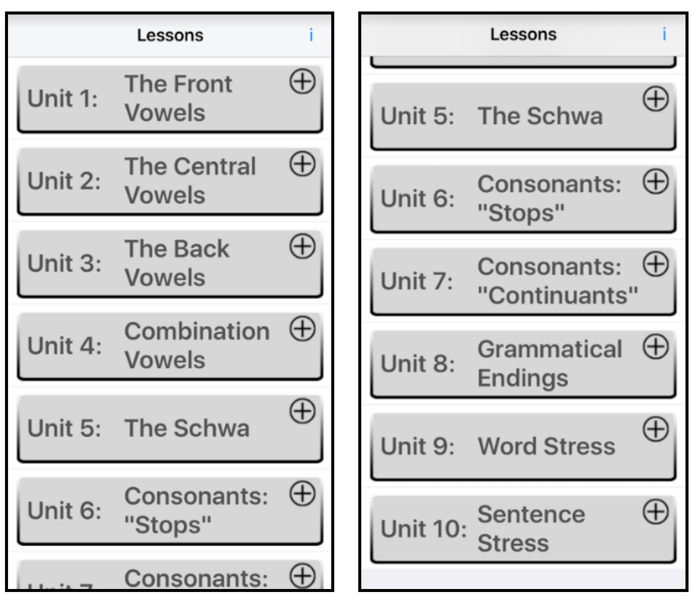


Figure 1. Units included in *American English Pronunciation Tutor*

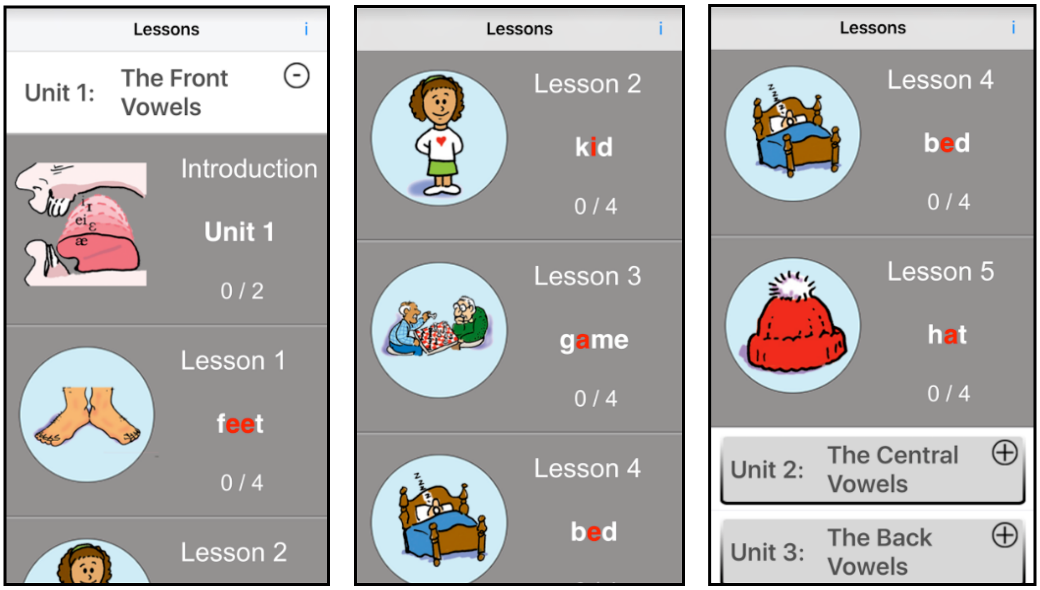


Figure 2. Lessons included in Unit 1

The Introduction section deals with articulatory aspects of pronunciation. In the case of Unit 1, there is an introduction to tongue height and mouth shape and how they are used to produce front vowels. Both the page for tongue height and the page for mouth shape have images that the user can click on to listen to audio files of the target vowels (Figure 3).

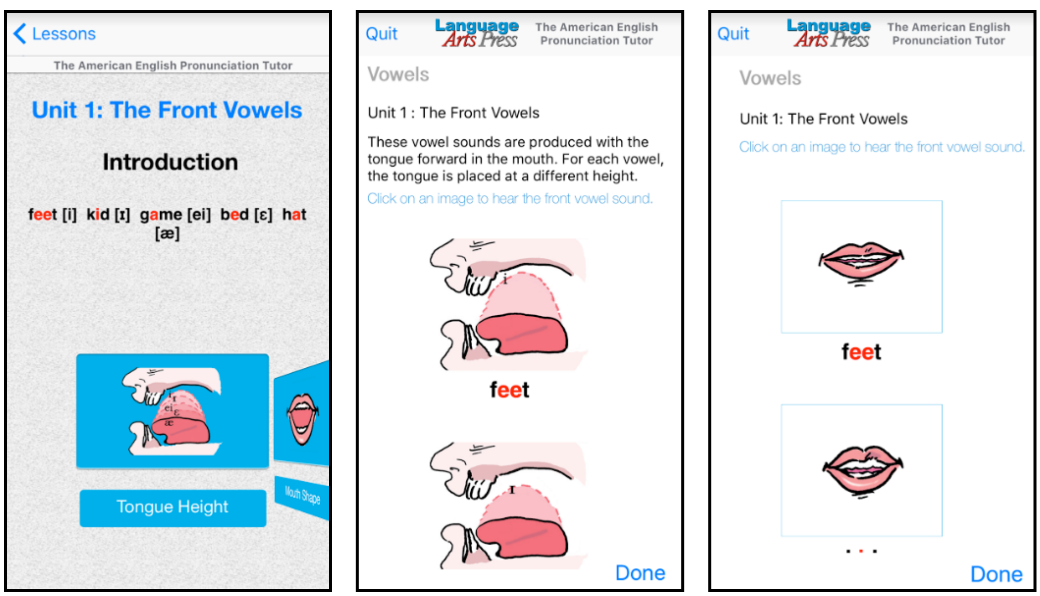


Figure 3. Introduction to Unit 1

Inside each lesson following the introduction, there are four exercises: Practice, Contrasts, Listening Quiz, and Speech Recognition. For the Practice exercise, the user listens to a word and records herself saying the word. Then, she can play the recording and compare it to the model (Figure 4).

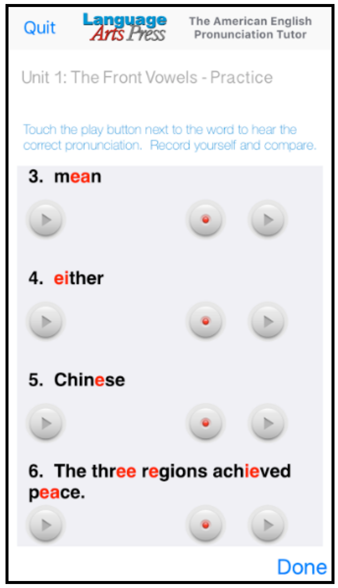


Figure 4. Practice exercise for Unit 1

For the Contrasts exercise, the user listens to minimal pairs. The same illustration for mouth shape is provided again for the user to see how the distinction is produced (Figure 5).

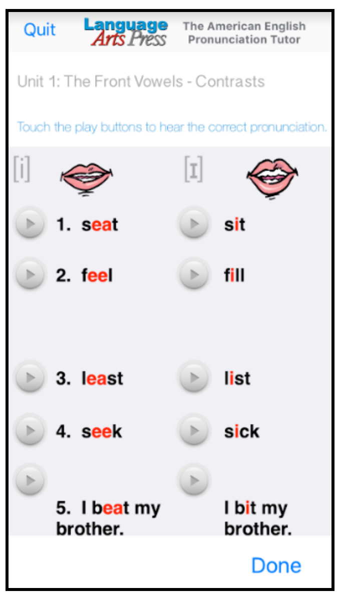


Figure 5. Contrasts exercise for Unit 1

For the Listening Quiz exercise, the user listens to a word and decides which of the two words was said. Immediate positive or negative feedback is provided. At the end of the Listening Quiz exercise, the user is prompted to attempt the exercise multiple times for additional practice with the other questions in the question bank (Figure 6).

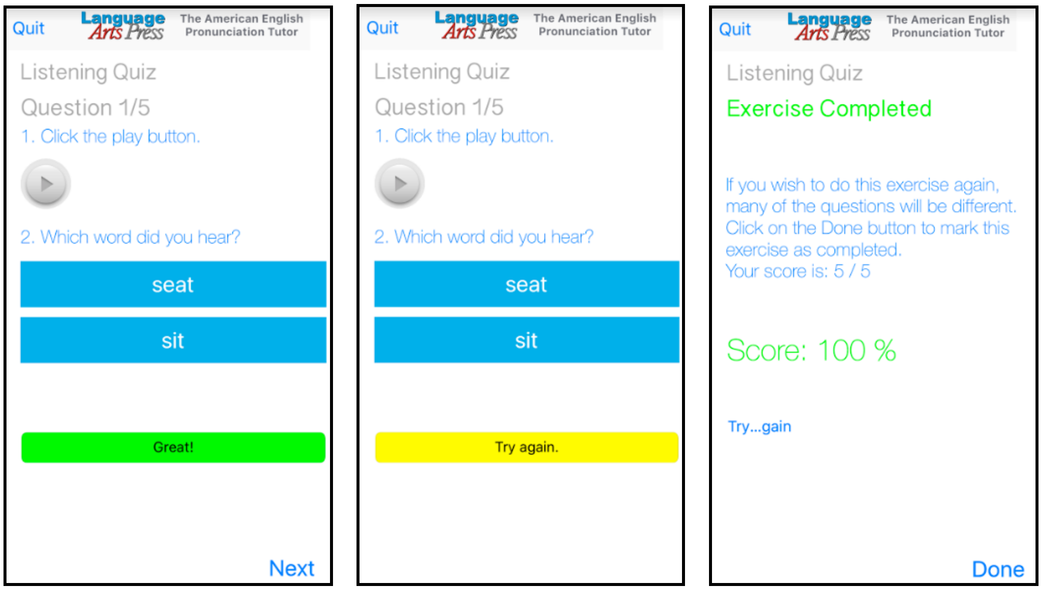


Figure 6. Listening Quiz exercise for Unit 1

For the Speech Recognition exercise, the user chooses a word from the Word Wheel, which includes words containing the target vowels. The user then records herself saying the word, and immediate positive or negative feedback is given, depending on whether the Automatic Speech Recognition system recognizes the recording as the target word (Figure 7).

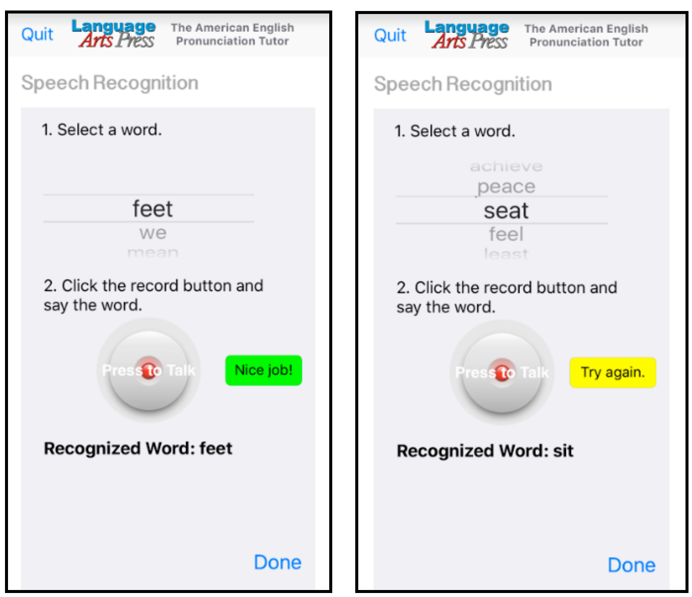


Figure 7. Speech Recognition exercise for Unit 1

The exercises logically progress from production practice for the target vowels to perception practice for minimal pairs of the target vowels, to perception assessment of those minimal pairs, and finally to production assessment of the target vowels. Since each lesson comes with all four exercises, Unit 1 comprises a total of 20 exercises on front vowels.

EVALUATION

The first thing to note about the *American English Pronunciation Tutor* is that it does not market itself as an app for accent reduction; rather, it focuses on the development of production and perception of segmental and suprasegmental features typical of speakers of American English. This focus on both segmentals and suprasegmentals has the potential to make this app more effective at training users for spontaneous speech, since Derwing, Munro, and Wiebe (1998) found that students who received pronunciation instruction that covered both segmentals and suprasegmentals demonstrated improvement in spontaneous speech abilities.

Another thing to note is that the *American English Pronunciation Tutor* app uses both orthography and International Phonetic Alphabet (IPA) symbols to represent the target vowels. Because English has an opaque orthography (Erdener & Burnham, 2005), almost every sound can be represented by different orthographic shapes, causing orthographic interference for speakers of languages with transparent orthographies like Spanish and Turkish, both of which have one-to-one sound-letter correspondence in pronunciation (Bayraktaroğlu, 2008; Erdener & Burnham, 2005). By including the different spellings for one vowel sound, the app can reduce orthographic interference by training users to recognize some of the common orthographic shapes for a particular vowel sound.

This may be challenging at first, especially for users who speak languages with transparent orthography and who are unfamiliar with the "marked lack of economy in the choice of letter representation" (Bayraktaroğlu, 2008, p. 2) in English, but Erdener and Burnham (2005) have shown that learning an opaque orthography can facilitate the efficient processing of inconsistent information like spelling and pronunciation in the long run, further reducing orthographic interference for these learners.

On the flipside, Saito (2012) pointed out that pronunciation activities should move away from controlled, drill exercises and towards meaning-oriented communicative exercises in order to be more effective at training students for improvement in spontaneous speech abilities. In the *American English Pronunciation Tutor* app, all four exercise types (Practice, Contrasts, Listening Quiz, and Speech Recognition) are controlled exercises aimed at eliciting production and perception of the target sounds only. If the app truly wants to help users develop clear, confident speech, it should consider incorporating less controlled or communicative tasks.

In line with the limited types of exercises, the *American English Pronunciation Tutor* app also falls into the category of edutainment (Gros, 2007), the genre of game that provides opportunities to practice the same skill multiple times with the assumption that the skill will be acquired, provided it was repeated enough times. Current theories on game-based learning, or the use of digital games for educational purposes, state that games should "provide complex environments in which content, skills, and attitudes play an important role during the game" (Gros, 2007, p. 26). In order to do so, the app should incorporate more game-based learning theories and either add to or adapt the existing four exercise types to reduce repetitiveness and increase user interaction.

SUMMARY

At first glance, the *American English Pronunciation Tutor* app has its merits. It focuses on the development of production and perception of the segmental and suprasegmental features of American English without advocating for accent reduction, and it provides orthographic representations of the target sounds along with the IPA symbols, which could be useful for some learners. However, the app is still lacking in the sense that it should incorporate more theories of game-based learning by reducing the repetitiveness of its exercises by adding to them or adapting them to have meaning-oriented communicative exercises.

REFERENCES

Bayraktaroğlu, S. (2008). Orthographic interference and the teaching of British pronunciation to Turkish learners. *Journal of Language and Linguistic Studies*, *4*(2), 1-36. Retrieved from http://dergipark.ulakbim.gov.tr/jlls/article/view/5000084222

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.doi:10.1111/0023-8333.00047

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. Retrieved from http://dx.doi.org/10.1111/j.0023-8333.2005.00303.x

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. Retrieved fromhttp://files.eric.ed.gov/fulltext/ej826060.pdf

Language Arts Digital. (2016). American English Pronunciation Tutor (Version 3.0) [Mobileapplication software]. Retrieved from http://itunes.apple.com

Language Arts Press Products. (nd.). Retrieved from http://languageartspress.com/products.html

Saito, K. (2012). Effects of instruction on L2 pronunciation development: A synthesis of 15 quasi-experimental intervention studies. *TESOL Quarterly*, *46*(4), 842-854. doi:10.1002/tesq.67

Silpachai, A. (2017). Accent reduction for professionals: How to eliminate your accent to sound more American (Review). 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. 270-273). Ames, IA: Iowa State University.

BOOK REVIEW

Nelson, W. (2015). *Accent Reduction For Professionals: How to Eliminate Your Accent to Sound More American*. Self-published. Kindle Edition. (110 pages in print)

Alif Silpachai, Iowa State University

INTRODUCTION

Just as scientists must educate the general public through science communication, pronunciation researchers too should bridge the gap between research and practice (see Levis, 2016 for current review). This was why I chose to investigate pronunciation teaching in non-academic settings. I found this book on Amazon.com. It was the first item that showed up after I typed "accent reduction." It has positive reviews although there are only sixteen reviewers. I purchased the Kindle version for $2.99 (which is a lot cheaper than the print version which is $34.79).

As the title suggests, the goal of this book is to help learners of English reduce and eliminate their foreign accents to be successful in their careers in the US. The author, Whitney Nelson, writes that accent elimination is easy, or in her words "Before you begin to think that it's impossible to eliminate your foreign accent, I'm here to tell you that it's not only possible, but probably easier than what you believe it to be" (171-172). This shocking claim goes against current research, which suggests that accent elimination is extremely difficult; it is instead more practical to train learners to become intelligible L2 speakers (e.g., Derwing & Munro, 2005; Jenkins, 2000; Munro & Derwing, 1995).

According to her Amazon profile, the author is an "international English professor" who has been teaching EFL courses for 23 years "at a few big organizations in Korea and China, Open Universities, technical colleges, and high schools." Her other books on the website are all about improving one's speaking skills. Unless she uses a pseudonym, I could not find her professional, university-affiliated web page, or any peer-reviewed journals authored by her. This frankly makes me skeptical about her expertise in pronunciation.

The introduction addresses a few questions concerning accent elimination, e.g., "What is an accent?" and "Is eliminating accent really necessary?" etc. She describes two types of accents: native accent (e.g., New York accent) and foreign accent. Her main rationale for accent elimination is based on the premise that NSs "have difficulty understanding their [an NNS's] strong accent" (146). Thus, her definition is different from Munro & Derwing (1995) who do not conflate accentedness with intelligibility and comprehensibility. Her definition also suggests that her teaching shares the nativeness view (Levis, 2005).

Chapter 1 attempts to answer "Is your strong accent preventing you from getting your perfect job?" She argues that it does, and to support her claim, she uses discrimination as the main reason. For example, she cites Dianne Markley, a professor at the University of

North Texas at Denton, who says that "there exists an 'incredibly strong statistical correlation between judging someone as cultured, intelligent, and competent and placing them into prestigious jobs' based on the lack of an accent" (242-244). Although discrimination against accents certainly exists in certain places, she still needs to provide evidence that clearly supports the claim that people's foreign accents prevent them from getting a "perfect job," let alone being "successful" in their careers. Also, her argument is flawed because she makes the assumption that NNSs' "perfect jobs" will be discriminatory. A person's idea of a "perfect job" might be a place where foreign accents are highly valued. For me, my "perfect job" would be a place where I am not discriminated against because of my accent.

In chapters 2-7, the author gives lessons on sound production. In chapter 2, she provides tips on how to master word stress, although sentence stress is briefly mentioned too. She rightly points that English stress is not always predictable and urges learners to listen to NSs as models for best results. Considering the book's aims, this chapter's lesson is appropriate for the audience. From what we know from the literature, accentedness in English is correlated with word stress (e.g., Crowther, Trofimovich, & Isaacs, 2016). Thus, focusing on this area might make learners sound less accented. However, the literature also informs us that accentedness is correlated with rhythm in English (ibid). She therefore should have also included a discussion on rhythm with the focus on vowel reduction.

In chapter 3, entitled "Learning the Sounds of the Vowels," the author draws readers' attention to only five vowels: a, e, i, o, and u (IPA: /eɪ/ /i/ /aɪ/ /oʊ/, and /u/ respectively). Although I can see why these canonical five vowels might be appealing to her readers, this chapter does not accomplish the book's goals. These vowels are sufficiently distinctive from each other. If a reader wants to be understood better, he/she must work on high functional load vowel pairs, e.g., /i/ and /ɪ/ (Brown, 1988). Another concern I have is the organization of the chapters. Chapter 7, entitled "Discovering Diphthongs," is also about vowels (two diphthongs: /ɔɪ/ and /aʊ/). Three of the vowels she discusses in chapter 2 are also diphthongs. Why not combine these two chapters?

Chapter 4 offers a discussion on the speech organ responsible for nasality — the soft palate, and its role in accent elimination. She notes that "the goal is to sound less nasal" (591). This goal is not useful because it makes the assumption that nasality hinders comprehension and intelligibility, and that workplaces discriminate against nasalized speech. It also assumes that her readers' English is nasalized and that there is no nasalization in English. To make her lesson more meaningful, she should change the goal of this chapter to "learn how to nasalize like a native speaker" (e.g., syllables with nasal codas; CVN).

In chapter 5, the author highlights the /ɹ/ sound. According to her, it "comes in *three varieties*: or, ar, and air… [and] to make this sound correctly, you'll move the tip of your tongue toward the back of your mouth, pointing it backwards and flex it." (619-620, emphasis added). Her description is misleading because these are clearly three vowel phonemes which may influence the phonetic realization of /ɹ/. Varieties of /ɹ/ targets have in fact been identified in the literature, and they can be broadly categorized into two types: bunched (tip-down) and retroflex (tip-up) (e.g., Mielke, Baker, & Archangeli, 2016). She seems to be describing the latter type. My other concern is that if learners want to be understood, they should prioritize learning initial /ɹ/. There are many non-rhotic varieties of American English (e.g., Boston English), and rhotic speakers do not seem to have difficulty understanding these non-rhotic varieties. Thus, to meet the book's goals, the author should have also included a lesson on initial /ɹ/.

Chapter 6 provides a lesson on /θ/ and /ð/, which are ones of the rarest sounds in the world's languages (Maddieson, 1984) and bear a low functional load (Brown, 1988). Although some researchers do not recommend prioritize teaching these sounds (e.g., Jenkins, 2000), considering that the book's main goal is to help readers master the mainstream American accent, the inclusion of this chapter is justified.

As mentioned above, chapter 7 is about two diphthongs: /ɔɪ/ and /aʊ/. I have already stated my view on this chapter. I will not comment any further. In chapter 8, entitled "Putting it All Together," she provides reasonable tips on how readers can practice their pronunciation. Their tips all rely on listening to NSs either in person through shadowing, or via television, radio, podcasts, and YouTube.

In chapter 9, the author proposes another accent reduction method called "Reverse Accent Mimicry." In this method, learners mimic someone who speaks the same L1 as them, and the goal is to "just repeat what they say, [and] mimic everything you notice in how they speak, down to their gestures" (885-886). Then, the learners have to "transition this mimicry into" English by mimicking NSs of English. Although she does not provide a rationale for this method, I think the purpose might be to raise the learners' awareness of their L1 pronunciation. If true, this would mean that knowing one's L1 pronunciation will facilitate the reduction of L1 sounds when speaking a second language. I genuinely find this method interesting, but I remain doubtful of its effectiveness. The author should have cited research studies that have assessed its role in pronunciation learning.

Very similar to chapter 8, chapter 10 offers more tips and facts (dubbed "secrets") on English pronunciation. For example, other tips include "Practice, practice, practice," listening to audiobooks, and recording oneself etc. A few facts include English has voiced and voiceless sounds, the letter s is sometimes pronounced as /z/, and sounds run into each other in fluent speech etc.

For a book published in 2015, I found it disappointing and misleading. Throughout, the author claims that if people in this country get rid of their non-American accents, they will have a higher chance of being successful in their careers (often dubbed "perfect" or "dream jobs"). This book has not sufficiently provided evidence to support this claim nor the various assumptions which the claim is built on. With regard to her pronunciation teaching, the author should base her arguments on what the literature says. Research should inform practice.

REFERENCES

Brown, A. (1988). Functional load and the teaching of pronunciation. *TESOL Quarterly*, 593–606.

Crowther, D., Trofimovich, P., & Isaacs, T. (2016). Linguistic dimensions of second language accent and comprehensibility. *Journal of Second Language Pronunciation*, *2*(2), 160–182.

Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly*, *39*(3), 379–397.

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

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

Levis, J. M. (2016). Research into practice: How research appears in pronunciation teaching materials. *Language Teaching*, *49*(3), 423–437.

Maddieson, I. (1984). *Patterns of Sounds*. Cambridge University Press, New York.

Mielke, J., Baker, A., & Archangeli, D. (2016). Individual-level contact limits phonological complexity: Evidence from bunched and retroflex /ɹ/. *Language*, *92*(1), 101–140.

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.

Nelson, W. (2015). *Accent Reduction For Professionals: How to Eliminate Your Accent to Sound More American* (Kindle Edi). Whitney Nelson (self-published).

Yamashita, T. (2017) Manythings.org (Review). 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. 274-280). Ames, IA: Iowa State University.

WEBSITE REVIEW

Manythings.org

Taichi Yamashita, Iowa State University

INTRODUCTION

ManyThings.org (http://www.manythings.org/) is designed for L2 learners of English to self-study several aspects of English outside classroom. It was developed by American ESL teachers at Aichi Institute of Technology in Japan, namely Charles Kelly and Lawrence Kelly. This teacher-made website explicitly states that it is intended for people learning English as a Second Language (ESL) and English as a Foreign Language (EFL). Also, it is non-commercial and free for users. The website provides multiple sections, such as *English Sentences Focusing on Words and Their Word Families*, *Jokes*, *Vocabulary Lists with Games and Puzzles*, *Matching Quizzes*, *Pronunciation*, *Daily Listen & Repeat*, *Grammar*, and *Listen & Read Along*. Each section prefers a certain device (e.g. mobile, tablet), and the website lists the preferable device for every section. According to the description, out of 100 sections available as of December 2016, 58 are promised to work in mobile, 38 in tablet, two in computer with Flash, two in computer with Java, and there are no sections that work in computer with RealAudio. The website says that those that work in mobile and tablet will work on standard computers. Among such variety, the platform that is reviewed in this paper is the section for pronunciation learning (http://www.manythings.org/pp/). Henceforth, the term website in this paper refers to the pronunciation section.

OVERVIEW

To begin with, the title being *American English Pronunciation Practice (For ESL/EFL)*, it is obvious that English adopted for the website is American English, though we cannot tell which varieties of American English are adopted. When you click on the link, you will see the page shown in Figure 1:

Figure 1. The main page of the pronunciation section

The section is further divided into sub-sections, namely *Minimal Pair Practice & Quizzes*, *How to Use These Pages*, *"Listen and Repeat" Videos*, *Songs & Poems*, and *Tongue Twisters*. Due to space limitation, this review introduces and discusses only *Minimal Pair Practice & Quizzes*.

This section consists of 24 lessons with a single minimal pair shown in each. Table 1 is the list of minimal pairs available:

Table 1

Minimal pairs available in the section.

Lesson	Minimal pair	Lesson	Minimal pair
1	teen – ty (e.g. 13 - 30)	13	made - mad
2	ferry - very	14	run - rung
3	late - let	15	look - luck
4	lake - rake	16	climb - crime
5	fond - found	17	hot - hat
6	these - Z's	18	they - day
7	said - sad	19	run - ran
8	best - vest	20	lugs - lungs
9	not - note	21	not - nut
10	thick - sick	22	thought - taught
11	bus - boss	23	eat - it
12	see - she	24	few - hue

For each minimal pair, learners can listen to a single word (but sentence for lesson 1) as many times as they like in *Practice*. In *Quiz*, when learners click on *Listen, Then Choose*, the audio is played randomly, and learners choose which word they think is played by clicking on the word icon in *Quiz* (see Figure 2):

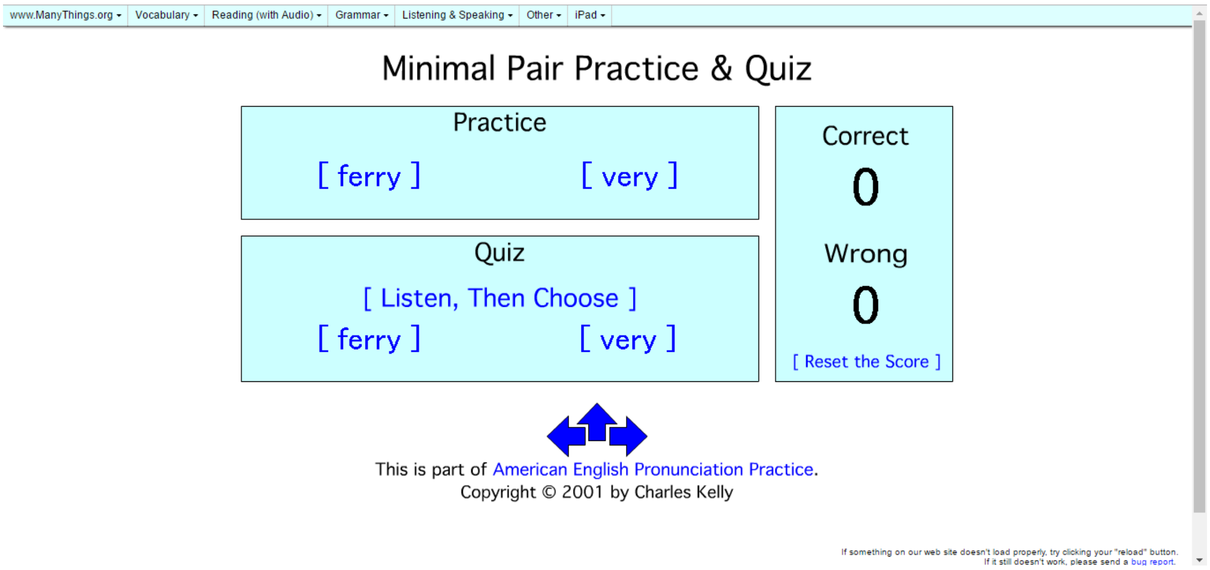


Figure 2. Minimal pair practice & quiz

If the user gets the correct answer, s/he gets one point under *Correct* (see Figure 3); otherwise, s/he gets one point under *Wrong* (see Figure 4):

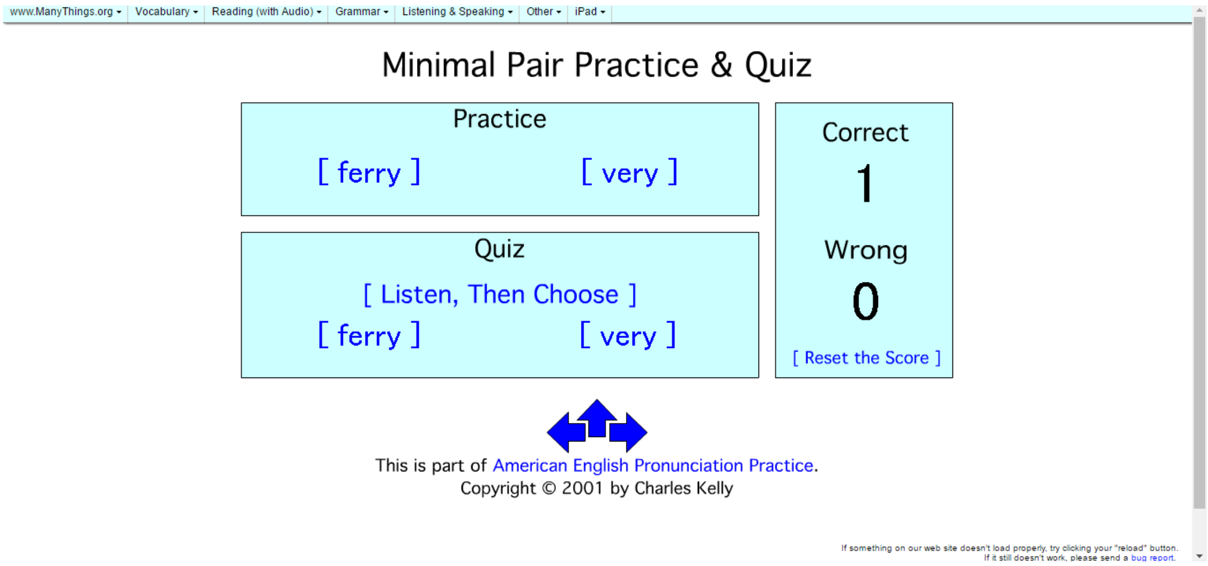


Figure 3. The page shown when the answer is correct

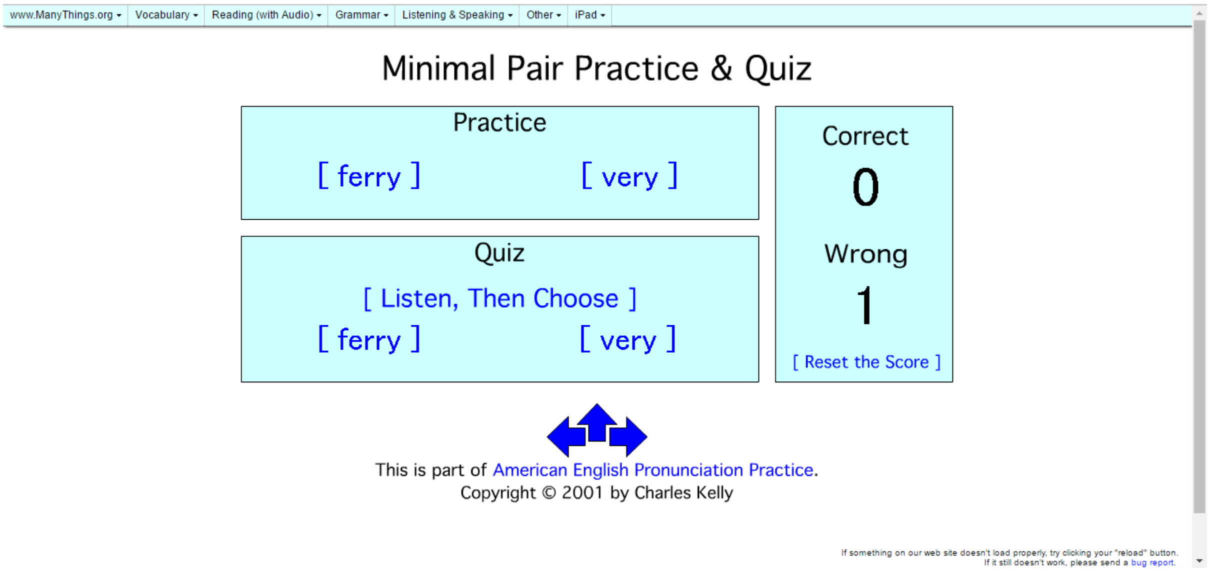


Figure 4. The page shown when the answer is wrong

EVALUATION

Following other research that has evaluated computer-assisted language learning (CALL), this review is also based on Chepelle's oft-cited work (Chapelle, 1997; Chapelle, 1998; Chapelle, 2001; Chapelle, 2003; Chapelle, 2005; Chapelle, 2009). One of her pioneering arguments is an appeal for approaches to CALL that are theoretically grounded in instructed SLA (Chapelle, 1997). Her argument is primarily supported by interactionist theory (Long, 1996), where the sequence of input, output, feedback, and modification drive L2 learning. From this perspective, the website may contribute to L2 learning but to a limited extent on the ground of the following pros and cons.

One of the pros about this website is the provision of feedback for learners (see Figures 3 and 4). Feedback is pivotal in SLA in that it provides what is possible in the target language (i.e. positive evidence) and what is not possible (i.e. negative evidence) for learners (Long, 1996). For this website, learners are given aural stimuli (input), demonstrate their current recognition of the stimuli, and receive either positive or negative evidence (feedback). Through this process, learners' interlanguage is gradually restructured and becomes closer to the target language.

The second pro is the way that minimal pairs are provided; learners can play a certain word in *Practice* repeatedly before working on *Quiz*, and, in both *Practice* and *Quiz*, each word is played in isolation. As for the benefit of repetition, Chapelle (2003), discussing vocabulary and grammar learning, indicated that it is beneficial for those who have some linguistic knowledge. In fact, a number of L2 pronunciation studies employ repetition of a certain word as instruction (Saito, 2012). Regarding isolation, it lets learners focus only on aural input without having them associate it with meaning. In fact, some studies have shown that learners of low proficiency have not automated sound-script and word-referent processing (Goh, 2000). Furthermore, since the answer is randomized, learners cannot rely on any patterns of answers, which is another encouraged operationalization of CALL (Nakata, 2011 for vocabulary learning).

Despite these pros, there are several cons as well. First, learners do not have an opportunity to produce a word. Therefore, though his argument is about morphosyntax, following DeKeyser's (1997) Skill Acquisiton Theory, the website is not as expected to improve learners' productive knowledge (i.e. automatized differentiation in production) as it is to improve their receptive knowledge (i.e. automatized differentiation in recognition) through the repetitive training. On the other hand, this recognition-centered nature is friendly for learners of low proficiency. Accordingly, the lack of production may not be satisfying for intermediate or advanced learners though it might be for beginning learners.

The second con is that the website is extremely mechanical where learners merely focus on aural input without paying attention to meaning or form. Though this factor was taken as an advantage in the previous discussion, recent studies in L2 pronunciation have criticized such mechanical drills in the form of repetition (Gooch, Saito, & Lyster, 2016; Saito, 2012; Saito & Lyster, 2012). However, such criticism originates mainly from the apprehension of teachers that such laboratory-like repetition is not authentic or applicable to classroom teaching where meaning should be the focus. Therefore, when learners use this website outside classroom for the purpose of self-study, this disadvantage might be mitigated. In conclusion, the website is not fully satisfying in terms of ideal conditions for SLA, but as long as learners utilize it merely for improving their receptive differentiation of minimal pairs, it would be beneficial for them.

REFERENCES

Chapelle, C. (1997). CALL in the year 2000: Still in search of research paradigms? *Language Learning & Technology*, *1*(1), 19-43.

Chapelle, C. A. (1998). Analysis of interaction sequences in computer-assisted language learning. *TESOL Quarterly*, *32*(4), 753-757.

Chapelle, C. (2001). *Computer applications in second language acquisition*. New York: Cambridge University Press.

Chapelle, C. (2003). The potential of technology for language learning. In Chapelle, C., *English language learning and technology* (pp. 35-68). Philadelphia, PA: John Benjamins Publishing.

Chapelle, C. A. (2005). Interactionist SLA theory in CALL research. In J. Egbert & G. Petrie (Eds.), *CALL research perspectives* (pp. 53-64). Mahwah, NJ: Lawrence Erlbaum.

Chapelle, C. A. (2009). The relationship between second language acquisition theory and Computer-Assisted language learning. *The Modern Language Journal*, *93*(1), 741-753.

DeKeyser, R. M. (1997). Beyond explicit rule learning: Automatizing second language morphosyntax. *Studies in Second Language Acquisition*, *19*, 195–221.

Goh, C. C. (2000). A cognitive perspective on language learners' listening comprehension problems. *System*, *28*(1), 55-75.

Gooch, R., Saito, K., & Lyster, R. (2016). Effects of recasts and prompts on L2 pronunciation development: Teaching English /r/ to Korean adult EFL learners. *System*, *60*, 117-127.

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. Ritchie & T. K Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413-468). Academic Press.

Nakata, T. (2011). Computer-assisted second language vocabulary learning in a paired-associate paradigm: A critical investigation of flashcard software. *Computer Assisted Language Learning*, *24*(1), 17-38.

Saito, K. (2012). Effects of instruction on L2 pronunciation development: A synthesis of 15 quasi-experimental intervention studies. *TESOL Quarterly*, *46*(4), 842-854.

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.