PROCEEDINGS OF THE 7TH ANNUAL PRONUNCIATION IN SECOND LANGUAGE LEARNING AND TEACHING CONFERENCE (ISSN 2380-9566)

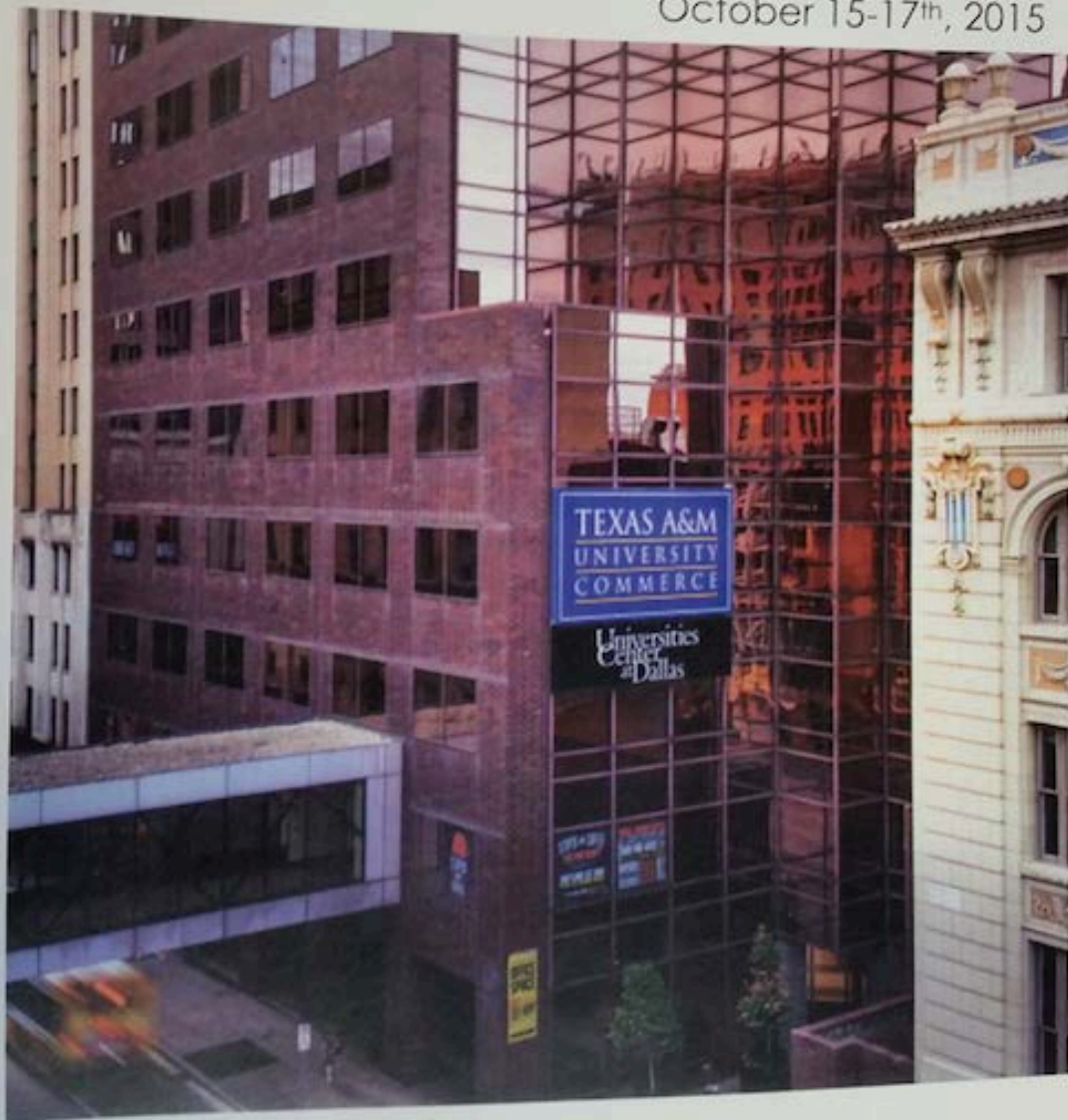## Editors

John Levis

Huong Le

Ivana Lucic

Evan Simpson

Sonca Vo

L2 PRONUNCIATION AND DISCOURSE

Pronunciation in Second Language Learning and
Teaching
PSLLT
7th Annual Conference
October 15-17th, 2015

PRONUNCIATION IN SECOND LANGUAGE LEARNING AND TEACHING
7th Annual Proceedings
(Conference held October 2016, Calgary, Alberta)

TABLE OF CONTENTS

Experimental and Instructional Approaches

Technology and L2 Pronunciation

Teaching Tips

Reviews

THE IMPORTANCE OF DEDICATED CONFERENCES
TO THE FIELD OF SECOND LANGUAGE PRONUNCIATION

John Levis, Huong Le, Ivana Lucic, Evan Simpson, Sonca Vo
Iowa State University

This is now the seventh year of the PSLLT Proceedings, with now approximately 150 published articles of various sorts, and almost 500 presentations. The conference started at Iowa State University and has also been held at Simon Fraser University in Vancouver, British Columbia, Santa Barbara, California, Dallas, Texas, and this year in Calgary, Alberta. Each year, the conference attracts 100-120 established and new researchers from 15+ countries.

The Pronunciation in Second Language Learning and Teaching conference has not only been relatively successful, more importantly it has provided a meeting place for pronunciation researchers and teachers working on a variety of languages from a wide variety of perspectives. As the number of second language pronunciation conferences increases, the PSLLT conference holds, in our view, a special place in the developing field of second language pronunciation. It is a conference that is not focused on English pronunciation, unlike other similar conferences. It has given birth both to these proceedings and to the new Journal of Second Language Pronunciation. So what is the conference like?

The conference takes part in late summer/early fall, lasts two full days, on a Friday and Saturday. Registration includes the conference, one lunch, snacks and drinks at all breaks, the conference dinner, and a chance to meet and talk with current and future big names in the field. There are a variety of presentation types, from oral talks to a dedicated poster session that does not compete with any other presentations, to a Teaching Tip Roundtable session, periodic special colloquia and sessions, as well as plenary speakers, and a pronunciation book give-aways from publishers at the end of the conference. One of the best things about the small size of the conference is that, unlike many other conferences, there are no barriers between the new and established participants. We asked conference attendees last year to give us feedback about the conference, and these are the themes that emerged from their feedback.

1. **The conference was enjoyable due to many factors.**
   a. Attendees enjoyed the structure and length of the conference. It was convenient having the poster presentations during lunch, and not during other sessions, so that there was time to experience everything.

"I thought the conference was fantastic! The presentations were well done, the teaching tips were informative, and the food and company was also excellent."

    b. Others liked the diversity of the presentations that were offered.

"I found the variety of pronunciation-focused presentations valuable; in particular, the balance between research-oriented and practitioner-based was great. I have been on the practitioner side for the last 15+ years (and my full-time teaching position doesn't require a research component), so I appreciate seeing what is being empirically investigated these days."

"Attending presentations which are very different than my research. For example, I found the first presentation about Forensic linguistics extremely interesting and mind opening."

    c. Many enjoyed the opportunity to network and meet new people in and outside of their field.

"I appreciated the opportunity to talk with both junior and senior researchers who have the same research interests that I do."

"I found the ability to network and talk about research to be the most valuable aspect."

"Also extremely valuable is the time allotted to networking in the schedule-- not just the reception and meals, but also the scheduled breaks. As a result, there are a few people that I might now collaborate with for future research and presentations."

    d. The intimacy of the conference was a nice change and allowed for more individual interactions.

"The best thing for me is that the conference continues to be intimate, but large enough to be quality. It's great to see how everyone is made a part of the PSLLT community. It really felt like more of a community this year than ever."

"I was pleasantly surprised as a first-time attendee to see how supportive and down to earth attendees are. Perhaps I was anticipating in-group / out-group tensions, or super-sized egos of well-published people--anyway, I experienced nothing of the sort. This seems to be a fine group of scholars who want the field to advance, and are happy to support each other in doing so."

"I really appreciated the size of the conference. I feel like I had the chance to talk to many people, especially during the poster session."

    e. Others highly appreciated the quality of the presentations at the conference.

"PSLLT was a high quality conference with a most congenial atmosphere. I think the high point of the conference for me was probably the two speech rhythm presentations on Saturday afternoon. Dr. Munro's research in this area relates directly to my own, and so I was very much

looking forward to his presentation. It did not disappoint. The subsequent presentation by Dr. Dickerson was one of the most succinctly organized I have ever heard. I should say, however, that every presentation I attended was very strong. These two happened to be my favorite because of my own personal interests."

    f.   Both experienced and novice researchers got benefits from good feedback on their work.
"people's comments, questions and reactions to my research. It gave me idea for further directions and gave me ideas about what to add to the future articles I intend to write."

"people sharing their research. I send some articles to people and some people sent me the articles they wrote about a similar topic. This way we can help each other and benefit from each other's research."

    g.   The location of the conference was also an important factor in the satisfactory experiences of many conference attendees.

"I was grateful that our hotel was within close proximity to the conference location."

2. Some of the best memories of the conference
    a.   involve the opportunities to interact openly with peers.

" I think the opportunities that are built in for ease of conversations - this
year's Thursday reception, the provision of the Friday lunch, the Friday dinner, the focal venue space - make it possible that this is where people can confer and set their research agendas for future collaborations and/or projects."

    b.   Networking opportunities with former colleagues and experts

"Reconnecting with colleagues and making new connections. Also, on a very personal level, this was the first time I came with my former student, who is on the TT. She was thrilled with the conference and I was so proud to introduce her to everyone."

    c.   Everyone loved the atmosphere of the conference and enjoyed the time that was left for open conversation.

"Experts and newbies were mixing and sharing ideas and excitement for what is happening in the field. The 'big name experts' did not stick exclusively together; they were open and amenable - even making it a point - to engaging in conversation with those they did not know."

"I like the atmosphere of this conference: professional, insightful and yet relaxed and friendly."

"I thoroughly enjoyed the conference. The highlights for me were (a) visiting with past students and seeing how well they were doing professionally and personally, and (b) making some new friends who share an interest in the work I'm doing. The conference has grown in size and also in quality of papers, which bode well the future."

2. **While the conference was well organized, a few small things that could help decrease general confusion.**

a. It would be useful to have the Teaching Tips online, so that they could be used as a reference later on. Also, with the Teaching Tips, it would have been useful to have a brief synopsis of what each tip was, so that individual could be sure to listen for the tips they have the most interest in.
b. Timing was also an issue; it would be ideal to have more time for questions at the end of presentations.
c. Suggestions for future conferences:
   -having an 'Ask the Experts 'forum for seeking suggestions for research proposals (like have those wishing to avail themselves of the experts' advice submit their proposals in advance
   -having "discussion round-tables on specific topics"
   - a PRAAT training workshop the day/afternoon before PSLLT starts (and other useful tools with a steep learning curve)"
   - locating the conference near more affordable and plentiful hotel rooms
   - schedule conferences in the summer
   - more ways to meet and get to know more people during the conference: (1) Before the conference dinner, a reception for at least an hour, with everybody standing, drinking something (if they want) before sitting anywhere and start eating & (2) a small reception with everybody "standing" after the meal; (3) for the lunch of the second day, provide a "recommended place" where the conference participants would meet if they want to.

CONFERENCE SCHEDULE:

Thursday, October 15th: Opening evening reception: 5-6.30pm. 4th floor gallery			
Friday, October 16th			
Time			
8.00-8.50	Registration on 2nd floor		
8.50-9.00	Welcome in room 315/316		
9.00-10.00	Plenary Address: Ann Wennerstrom, ESL in handcuffs: Pronunciation and forensic linguistics.		
10.00-10.25	Break		
Morning Sessions	Room 315/316	Room 321	Room 307

10.30-10.55	**Zielinski, Beth. Wang, Jihong. Pryor, Elizabeth.** English use in everyday life: Is it important for the development of comprehensibility and fluency?	**Gomes, Maria Lucia.** Brazilian English x Brazilian Portuguese: A dynamic approach for the analysis of diphthongs in forensic contexts.	**Levis, John. Sonsaat, Sinem.** Pronunciation in the CLT era.
11.00-11.25	**Kang, Okim. Moran, Meghan. Thomson, Ron.** Measures of intelligibility in different varieties of English.	**Nagle, Charles.** Modeling the initial stages of pronunciation development: An investigation of L2 Spanish stops.	**Baker, Amanda. Burri, Michael. Acton, William.** Haptic instruction and L2 fluency development.
11.30-11.55	**Koffi, Ettien**. The acoustic phonetics of Eth in seven varieties of L2-accented English: Focus on Intelligibility.	**Bouchard, Julie.** French Canadian EFL speakers' prosodic orientation in (dis)agreement in French and English.	**Zhou, Ziwei**. A hip hop-based proposal to EFL pronunciation instruction: Bridging musicality and criticality.
12.00-1.50	**Lunch on 2nd floor**		

12.00-1.50	**Poster Session: Room 317 & 3rd floor lobby**
	Abat, Martina. Coda devoicing in western south Slavic speakers' accented English.
	Arnold, Erik. Smith, Laura. Baird, Kyle. Lau, Darrel. The effect of language experience on learners' perceptions of German vowels.
	Becker, Shannon. Improving perception of L2 French nasal vowels through high variability phonetic training.
	Carreno Galdame, Sofia Laura. Henrichsen, Lynn. Baker-Smemoe, Wendy. Tanner, Mark. A motivational online guide to help second language learners develop, implement and evaluate their individual pronunciation improvement plans.
	Cha, Jihyeon. Effects of Pitch adjustment on Pronunciation Correction
	Chan, Queenie. Munro, Murray. Processing time variability in foreign accent comprehension. **Crabtree, Janay**. In other people's words: Nonnative speakers' imitation of professional speech. **Divita, Sam.** Using adapted readers' theatre to improve young adult ELs' pronunciation of thought groups.
	Godfroid, Aline. Ryu, Catherine. Lin, Chin-Hsi. Colorful benefits: The efficacy of dual coding in an online L2 Chinese tone perception study.
	Gordon, Joshua. L2 pronunciation and classroom discourse: Teacher centered vs learner centered instruction.
	Huang, Meichan. Pickering Lucy. The pronunciation of English by speakers from a southern province in mainland China.
	Kermad, Alyssa. A study of NNS' comprehension of intonational meaning, in light of hours of TV/movies watched in English.
	Kinoshita, Naoko. The acquisition of Japanese rhythm: Is it lexical or rule-based?
	Lai, Wience Wingsze. Ng, Manwa Lawrence. A comparison between native English speakers' and Cantonese ESL Learners' English word stress perception.
	Lawson, Lynee. Letting the students speak: Lessons learned to maximize the effectiveness of peer feedback for oral presentations.
	Lee, Heeju. Prosody-syntax mismatches for holding turns: A study of English speaking Korean L2 learners.
	Noguchi, Masaki. Yamane, Noriko. Tsuda, Asami. Kazama, Misuzu. Kim, Bosung. Gick, Bryan. Towards protocols for L2 pronunciation training using ultrasound imaging.
	O'Neill, Sarah. Shea, Christine. Changes to self-correction following explicit pronunciation instruction.
	Rohr, Jessica. Kilpatrick, Cynthia. Story retelling and prosodic behavior.
	Sonsaat, Sinem. The role of teachers' books in pronunciation teaching: An Answer key or a complete guide?
	Stenseth, Jennifer. Guinn-Collins, Shannon. Intensive pronunciation clinic: Enhancing pronunciation instruction with speech language pathology.
	Sturm, Jessica. Grim, Frederique. Where does pronunciation stand in the 21st century foreign language classroom? Educators' and learners' views.
	Talley, Jim. What makes a Bostonian sound Bostonian and a Texan sound Texan?
	Zetterholm, Elisabeth. Tronnier, Mechtild. Recognition of final consonants by L2 learners.
	Valenzuela, Maria Gabriela. Comparative acoustic analysis of English vowels between Chilean Spanish and speakers of American English.
	Wallace, Lara. Technology use in pronunciation teaching: Current practices and hidden gems.
	Zetterholm, Elisabeth. Haslam, Mara. The importance of aspirated initial stops in English as a lingua Franca.

Afternoon Sessions	Room 315/316	Room 321	Room 307
2.00-2.25	**Thomson, Ron. Derwing, Tracey.** Is phonemic training using nonce words or real words more effective?	**Sardegna, Veronica. McGregor, Alison**. Changes in ESL Oral Proficiency after instruction: Read-aloud vs. Extemporaneous speech.	**Violin-Wigent, Anne**. I want to sound just like that." Student attitudes towards native and non-native models.
2.20-2.55	**Gess, Randall.** Exploiting corpus data in L2 pronunciation teaching: The phonology of contemporary French project.	**Foote, Jennifer.** The impact of shadowing on improving pronunciation in extemporaneous speech.	**Sakai, Mari.** Production training in the absence of sound.
3.00-3.25	**Zielinski, Beth. McGregor, Alison. Reed, Marnie. Meyers, Colleen**. In search of a teachable model of intonation: A perceptual, acoustic and interpretive investigation.	**White, Donald. Chan, Jason. Mok, Peggy. Lie, Peggy.** Mimic Video: A Cinematic method for L2 pronunciation instruction.	No session
3.30-3.55	Break		
4.00-4.25	**Munro, Murray. Derwing, Tracey. Halcro, Leeandria.** Longitudinal acquisition of rhythm in L2 English.	**Ben Abda, Imen.** The production of English prosody by native speakers of Tunisian Arabic.	**West, Richard. Wallace, Lara.** Beyond vowel and consonant charts: Identifying areas for improvement in discourse-level pronunciation.
4.30-4.55	**Dickerson, Wayne**. A practitioner's guide to English rhythm.	**Isiaka, Lasisi Adeiza.** Ebira and Yoruba English accents: a sociophonetic study.	**O'Brien, Mary. Dressler, Anja.** Assessing fluency vs. fluidity in L2 German Speech.
	Conference Dinner, Hotel Indigo		

Saturday October 17th	
8.00-8.50	Registration on 2nd floor
8.50-8.55	Announcements Room 315/316

| 9.00-10.30 | Teaching Tips Room 321 | **Chan, Marsha.** Improving stress and rhythm with the stress stretch.
De Moras, Nadine. Learning L2 pronunciation (French obligatory liaisons) while studying vocabulary.
Henrichsen, Lynn. Peer-tutoring pronunciation contrasts: A fun, effective classroom procedure.
Meyers, Colleen. The straw technique: Expanding pitch range.
Muller Levis, Greta.
Levis, John. Intonation bridging activities: Meaningful practice for final intonation.
Nibert, Holly. Bringing L2 classroom pronunciation practice in line with CLT.
Reed, Marnie. Teaching talk and tell-backs: The declarative to procedural knowledge interface.
Richards, Monica. Transforming any text into an individualized segmental exercise via the pronunciation highlighter.
Ruellot, Viviane. French pronunciation and vowel tension.
Wallace, Lara. Lima, Edna. Five winning activities for SPEAK test preparation.
Zhuang Yuan. Staples, Shelley. Using PRAAT to visualize suprasegmentals for language learners. |
| | | **Watts, Patricia. Lawson, Lynee.** Was that a question? Applying the noticing- the-gap to help speakers recognize and use phonological features.
Zhuang Yuan. Staples, Shelley. Using PRAAT to visualize suprasegmentals for language learners. |

10.30-10.55	Break		
Morning sessions	Room 315/316	Room 321	Room 307
11.00-11.25	**Jiang, Yan. Chun, Dorothy.** Individualized intonation training with visualization feedback.	**McCrocklin, Shannon,** The effectiveness of ASR- based dictation practice for pronunciation improvement.	**Hardison, Debra.** Communication strategies and oral interaction abilities in ESL learners: Role of interlocutor type.
11.30-11.55	**Zárate-Sández, Germán.** How is intonation in a second language perceived? The case of pitch alignment in Spanish.	**Durham, Kristie, Hayes-Harb, Rachel. Barrios, Shannon.** The influence of various visual input types in second languages learners' memory for the phonological forms of newly-learned words.	**Crowther, Dustin. Trofimovich, Pavel. Isaacs, Talia.** The perception of L2 English speech by nonnative listeners: The effect of L1 background.
12.00-1.25	**Lunch**		
Afternoon sessions	Room 315/316	Room 321	Room 307
1.30-1.55	**Harada, Tetsuo.** Factors affecting phonemic discrimination by early and late EFL learners in Japan.	**Reed, Marnie. Lacroix, Jennifer.** Metacognitive strategy instruction Improves L2 skills in processing aural input.	No session
2.00-2.25	**Johnson, David. Kang, Okim. Ghanem, Romy.** Language proficiency ratings: Human versus machine.	**De Moras, Nadine. Peguret, Muriel.** Does an early start and longer practice make perfect?	No session
2:25-2:55	**Break**		
3.00-3.25	**Chan, Marsha. Brinton, Donna.** What's hot 2015 – Insights from pronunciation practitioners.	**Levis, John. Muller Levis, Greta.** Spoken parentheticals in instructional discourse: Implications for ESP pronunciation instruction.	No session
3.30-3.55	**Ma, Judy. Henrichsen, Lynn. Cox, Troy. Tanner, Mark.** The role of pronunciation in second language speaking test ratings.	**Wallace, Lara.** Using Google Web Speech as a springboard for identifying potential pronunciation problems.	No session
4.00-5.30	Room 315/316: Update from John Levis on the *Journal of Second Language Pronunciation* Roundtable discussion/Q&A with members of the board of the journal.		

The Proceedings

This year we have 15 full-length papers, 8 Teaching Tips, and 10 Book, software, app and website reviews. We have divided them into sections: Experimental and Instructional Approaches (7), Technology and L2 Pronunciation (8), Teaching Tips (8), and Reviews (10). The reviews were not part of the conference but were done by graduate students at Iowa State University. A summary of each paper is included below.

Experimental and Instructional Approaches

In "What's Hot 2015: Insights from Pronunciation Practitioners", Marsha Chan and Donna Brinton investigate what kinds of topic that international pronunciation specialists elected to discuss over the one-year period from August 2014 to August 2015. They analyzed the e-list discussion strands and threads and showed four hot topics discussed the most: techniques for helping Vietnamese speakers learn English pronunciation; stress shifting in British and American English; the respective merits of differing vowel charts; and the value of contrastive analysis for research and teaching.

Janae Crabtree investigates international graduate students' enjoyment and perception of improvement due to using TED Talks voiceover as a pronunciation-improvement tool in an academic communication course ("In other people's words: Nonnative speakers' imitation of professional speech"). The author found that, even though participants improved their pronunciation and enjoyed the activity, they were not satisfied with their overall oral proficiency. Crabtree explores how perceived ability, raising awareness of learner's needs, and learner's identities connect and influence comprehensibility.

In "A practitioner's guide to English rhythm: a return to confidence", Wayne Dickerson traces the history and development of the stress-timed rhythm idea and its understanding in the field of TESOL, with intent to make it easier for ESL/EFL instructors to describe and teach English rhythm. He is also cautioning current and future practitioners about using the TESOL's model of rhythm in order to benefit both ESL/EFL instructors and ESL/EFL learners.

Frédérique Grim and Jessica Sturm argue that tertiary L2 learners hold pronunciation in higher esteem than do educators in "Where does pronunciation stand in the 21st century foreign language classroom? Educators' and learners' views." The authors make recommendations for future studies as for changes in pedagogical practices.

In "The importance of aspirated initial stops in English as a lingua franca," Mara Haslam and Elisabeth Zetterholm report an experimental study that examines the LFC's claim on the importance of the fortis-lenis contrast in ELF speech intelligibility. The study provides mixed results that both partially support and refute the claim. The authors then support a more granular approach to the LFC and give implications for further research on this issue.

10

Jennifer Lacroix, Marnie Reed and Allen Harbaugh, in "Metacognitive strategy instruction improves L2 skill in processing aural input", investigate the effect of strategy-based instruction on adult learners' beliefs and skills in aural input processing. A semester-long project showed that strategy-based metacognitive training in connected speech, stress and intonation promotes listening skills awareness, aids word segmentation to facilitate understanding utterance context, and helps detection of marked intonation to facilitate understanding of message meaning.

In "Is phonemic training using nonsense or real words more effective?" Ron Thomson and Tracey Derwing present the results of their exploratory study which sought to determine which method was most effective at teaching English vowels: presenting vowels in "isolated open syllables" or in "real words." The paper concludes by arguing for a greater focus teaching segmentals as well as devoting more time for practicing producing these sounds.

Technology and L2 Pronunciation

In "The Influence of Various Visual Input Types on L2 Learners' memory for the phonological forms of newly-learned words", Kristie Durham, Rachel Hayes-Harb, Shannon Barrios, and Catherine E. Showalter investigated whether the word form learning benefit reported in Showalter and Hayes-Harb (2013) is necessarily orthographic. Text position was found to provide a benefit over other non-orthographic visual information (color), as well as orthographic information (tone marks). The authors suggest that orthography, while a likely contributor to a performance benefit, is not the only beneficial visual information during word learning. The authors also call for studies that investigate the benefit of other types of visual information.

Setting in forensic contexts, Maria Lucia de Castro Gomes presents preliminary results of an experiment using acoustic analysis as a part of a project that is purposed to investigate special characteristics of the pronunciation of Brazilian speakers of English ("Brazilian English x Brazilian Portuguese: A Dynamic Approach for the Analysis of Diphthongs in Forensic Context"). Diphthongs are chosen to examine and compare the recordings of Brazilians and Americans in English and Portuguese in terms of inter-speaker, intra-speaker and inter-language. The author supports the hypothesis that Brazilian speakers of English may have special features that might indicate their origin, and describes the current work of the project.

In "Language Proficiency Ratings: Human vs. Machine," David Johnson, Okim Kang, and Romy Ghanem presents the development and evaluation of a computer model that automatically scores the English proficiency of unconstrained speech. Their automatic scoring system is reported to outperform other similar systems, and its scoring reliability is as close as inter-rater reliability of human scoring. Based on their results, the authors also imply the most important factors and potential factors for developing and improving such an automated system.

In "The acoustic phonetics of voiced TH in seven varieties of L2-accented English: Focus on intelligibility", Ettien Koffi analyzed the voiced interdental non-sibilant fricative [ð] produced by 10 native speakers of General American English (GAE) and 67 non-native speakers of English quantitatively and acoustically. The quantitative data shows that GAE talkers produced [ð] more

accurately than L2 talkers, substituted less frequently than L2 talkers. He showed that all the substitutions occurred only in syllable onsets, but not intervocalically. He argues that the substitutions do not compromise intelligibility because the relative functional loads between [ð] and the substitutions are very low, even negligible.

"An exploration of teaching intonation using a TED Talk" (Alison McGregor, Beth Zielinski, Colleen Meyers & Marnie Reed) used a multi-layered approach to investigate a TED Talk. The authors conducted an interpretative, perceptual and acoustic analysis of the monologic speech sample of North American English. The results show an integrated view of language and intonation, and support the idea of contextualized teaching and pedagogical use of TED Talk speech samples.

In "Transforming any text into an individualized segmental exercise via RelateWorldwide's *Pronunciation Highlighter*," Monica Richards introduces her online Pronunciation Highlighter, a tool that can transform any text that students find interesting such as textbook dialogues, technical term lists, PowerPoint presentation outlines into individualized segmental practice exercises. The author also suggests ways that students can capitalize on Pronunciation Highlighter output to build new and accurate segmental pronunciation habits.

Jim Talley presents the preliminary findings of a report on a "new data-driven methodology" in "What makes a Bostonian sound Bostonian and a Texan sound Texan?" The author argues that with time and refinement, this methodology could lead to an exhaustive catalog of the acoustic features that define accented speech.

In "Using google web search as a springboard for identifying personal pronunciation problems," Lara Wallace describes how and why L2 English learners should use automatic speech recognition (ASR) tools in order to improve their intelligibility. The paper includes suggested pedagogical practices as well as both the benefits and limitations of these practices.

Teaching Tips

In Marsha J. Chan's "The Stress of Prosodic Improvement in English words and phrases", she describes how to use the stress stretch with students learning to speak English. The stress stretch allows students to associate a physical movement to the concept of stressed and unstressed syllables to improve their pronunciation. Students stretch in accordance with the lexical stress or prominence of target words.

"French Pronunciation and Vowel Tension" by Viviane Ruellot examines the importance vowel tension plays in the comprehension of French Speakers. This tip emphasizes the role of vowel tension when discriminating between masculine and feminine forms of definite articles of nouns, and with third person singular direct object pronoun.

Lara Wallace and Edna F. Lima collaborated to write the article "Intelligibility: Five Winning Activities for Speak Test Preparation". Within the article, they present several activities to help International Teaching Assistants be more successful when they take the SPEAK Test. These

12

activities include fly-swatting fillers, cell phone persuasion, body language for better intonation, audacity and rhythm, and giving directions.

In Nadine de Moras' "Learning L2 Pronunciation While Studying Vocabulary", she describes syllabification in French. The tip focuses on how to help speakers learning how to speak French with liaison rules to help with pronunciation. These liaison errors are critical in that they impede understanding through a lack of differentiation between homonyms and impeding comprehensibility due to the missing connections.

"Peer-tutoring Pronunciation Contrasts: A Fun, Effective Classroom Procedure" by Lynn Henrichsen describes how this peer-tutoring sequence thrives on students different ability levels concerning the ability to discriminate between English segmentals and suprasegmentals. This peer-tutoring procedure places one high ability student with another lower ability student, concerning a specific topic, in the same group so that they may learn from one anther in both speaking and listening activities.

Greta Muller Levis and John Levis provide ideas for pronunciation bridging activities to practice English intonation. Bridging activities are between controlled and communicative activities, offering learners a chance to focus on pronunciation form but not completely, while paying attention to meaning, but not at the expense of accuracy. They show four ways to modify dialogues to practice intonation in ways that provide practice demanding attention to both accuracy and fluency.

In Marnie Reed's "Teaching Talk, Tell Backs, and a Declarative to Procedural Knowledge Interface" she describes how to use metacognitive coaching to link explicit and implicit knowledge gaps. English Language Learners often have language gaps that they are not aware of and do not know how to fix, but through provided metalinguistic feedback such as Teaching Talk, student Tell Backs, and Pronunciation Coaching, they can begin to make improvements.

"Providing individualized Homework and Accountability for ITAs via Internet Resources" by Monica Richards describes how it is important to implement the individual feedback on their spoken English that students receive emphasizing fluency, surprasegement al and segmental challenges. She makes the argument that individualized homework is beneficial and easier than ever to utilize through the resources available on the web.

Reviews

A supplement to the Proceedings this year is a set of reviews of pronunciation books, software, websites, and apps. These were written in John's graduate class on Technology and Oral Language class at Iowa State University in Fall 2016. They are included here to provide them a wider venue. I found the types of sites and apps and books chosen by the students interesting, especially because many were not items commonly spoken of in pronunciation circles, and knowing of them might be useful to other readers of the proceedings.

Chan, M. J., & Brinton, D. M. (2016). What's hot 2015: Insights from pronunciation practitioners. 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. 14-28). Ames, IA: Iowa State University.

WHAT'S HOT 2015: INSIGHTS FROM PRONUNCIATION PRACTITIONERS

Marsha J. Chan, Sunburst Media and Mission College, CA
Donna M. Brinton, Educational Consultant, Beverly Hills, CA

What do pronunciation specialists consider to be topics worthy of discussion amongst themselves? As a follow-up to previous studies of "hot topics" on an invitational electronic mailing list (e-list) for pronunciation specialists, this study investigates the issues that international pronunciation specialists elected to discuss during a one-year period. The authors, both members of the e-list, analyzed the e-list discussion strands and threads over the one-year period from August 2014 to August 2015 to determine the four topics that elicited greatest degree of interest, interaction, and in-depth discussion. The hot topics of this year, summarized here, are: 1) techniques for helping Vietnamese speakers learn English pronunciation; 2) stress shifting in British and American English; 3) the respective merits of differing vowel charts; and 4) the value of contrastive analysis for research and teaching.

INTRODUCTION

What do pronunciation specialists consider to be topics worthy of discussion amongst themselves? As a follow-up to previous studies of "hot topics" on an invitational electronic mailing list (e-list) for pronunciation specialists (Brinton & Chan, 2015; Brinton & Goodwin, 2006), this study investigates the issues that international pronunciation specialists elected to discuss during the period August 2014 to August 2015.[1]

To provide a flavor of the type and range of issues discussed on the e-list, we provide the following list of discussion strands, all of which generated healthy interest during the time period in question:

1. When learning an L2 with audio input, is written text a help or a hindrance?
2. To what extent does pronunciation correlate with overall language proficiency?
3. Of what importance is the vowel length distinction?
4. Is it possible for a pronunciation instructor to teach a better pronunciation than his or her own?[2]
5. Can a particular music genre (e.g., jazz) assist in pronunciation learning?

---

[1] At the time this article was written, the e-list consisted of 182 specialists from 25 countries around the world.

[2] This question referred to whether pronunciation instructors whose first language was colored by a certain dialect could teach the "standard" variant of the language.

Although these topics proved popular based on the number of responses and respondents, we chose not to analyze them due to a mismatch with our selection criteria (see below).

Integral to the invitational e-list discussion is the liberty to discuss issues large and small that pertain to pronunciation research and teaching, and to either participate actively (i.e., by posting questions and responding to queries) or passively (i.e., by "lurking" in an effort to inform oneself about the issues under discussion). As members of the invitational e-list, our goal in this study is to share highlights of those topics that elicited high interest and in-depth discussion over the one-year period. As Brinton and Chan (2015) note, there is great value in the e-list discussion format since "such a forum enables [specialists] to compare, challenge, debate, change and/or confirm ideas" (p. 161). In this article, we share these ideas in the interest of informing and enlightening the larger cohort of those interested in the field of pronunciation teaching and research.

Research Questions

For the purpose of this study, we were interested in pursuing the following questions:

1. On an invitational e-list discussion amongst English language pronunciation specialists, which topics are of current interest?
2. Of these, which topics elicited the greatest amount of response and in-depth discussion from the pronunciation specialists?

METHODS

In a previous study (Brinton & Chan, 2015) we analyzed discussion strands from the one-year period August 2013 to August 2014, narrowing our selection of topics to analyze based on those that had the greatest number of discussants and the largest number of exchanges. For the current study, our selection criteria included the following:

1. Number of words in the discussion threads
2. Number of exchanges
3. Number of discussants
4. General or global interest*
5. Depth of discussion*

Of these, the last two criteria (marked with an asterisk) are new to this study, and were added post facto after analyzing several strands that we believed, as specialists in the field, did not fit our own criteria (1) as being of sufficient interest[3] or (2) as having adequate depth of discussion to warrant analysis.[4]

---

[3] As an example, the August 2014-August 2015 exchanges included a rather lengthy discussion of the two possible pronunciations of the place name *Shrewsbury*. We eliminated this from our analysis based on our perception that the topic, while of interest to some, would not be of enough interest for the general reading public.

[4] This was the case with the discussion on the correlation between the attainment of pronunciation skills and general language proficiency. So while the question itself was provocative and queried participants as to any studies that

ANALYSIS

Applying the above criteria, we identified the following four discussion strands:[5]

1. How can we help Vietnamese speakers acquire a "listener friendly" pronunciation?
2. Do speakers of British English and American English shift stress differently?
3. What is your reaction to the proposed new pedagogical vowel chart of English which represents vowel length in concentric rings and vowel quality in radiating spokes?
4. Of what value is the contrastive analysis hypothesis to pronunciation research and/or teaching today?

Once the topics had been identified, we then downloaded the discussants' comments into separate documents to facilitate our task in compiling the data and summarizing the main ideas; we also created a separate document to capture the references shared by the discussants on the four topics (see Appendix).

As shown in Table 1, there is considerable variation in the number of words, discussants, and exchanges among the topics that we chose to analyze. The number of total words of the four strands varies from 5042 (Vietnamese speakers) to 1007 (contrastive analysis hypothesis, while the number of discussants ranges from 13 (Vietnamese speakers and vowel charts) to 8 (contrastive analysis hypothesis). As for number of exchanges, the strands range from 29 on the high end (vowel charts) to 9 on the low end (contrastive analysis hypothesis). A clear outlier among the topics is the contrastive analysis hypothesis, which exhibited the lowest value in terms of number of words, discussants, and exchanges. As noted above, this topic was selected on the basis of the additional criteria of general/global interest and depth of discussion, both of which the exchanges exhibited to a higher degree than alternative topics we could have selected such as that of vowel length distinction or the correlation of pronunciation skills with overall language proficiency.[6] The synthesis of each topic follows.

Table 1

E-list Topics Analyzed According to the Selection Criteria

	# of Words	# of Discussants	# of Exchanges
Vietnamese speakers	5042	13	23
Stress shifting in British and American English	2245	12	25

---

supported or refuted the correlation between the two, respondents tended to simply state their own opinions on the importance of learners acquiring pronunciation skills and not delve more deeply into the topic or cite related research.

[5] The questions have been slightly rephrased to clarify the author's original intent.

[6] These two topics were clearly of global/general interest. However, they were rejected due to their failure to fulfill the criteria of depth of discussion.

Vowel charts	3401	13	29
The contrastive analysis hypothesis	1007	8	9

Topic 1: Vietnamese Speakers of English

This thread was initiated when a practitioner teaching at university level queried the community on how to help Vietnamese speakers acquire a "listener friendly" pronunciation. The discussant, who hadn't worked with many Vietnamese speakers, referred to a "prevailing wisdom" of focusing on segmentals and word stress as well as addressing the airstream mechanism and using songs. Although one respondent opined that most of the Vietnamese speakers' segmental problems are phonotactically or syllabically based and downplayed prosody as secondary, several discussants with considerable experience teaching Vietnamese learners of English emphasized other aspects as being of equal or greater importance. Our analysis revealed three general categories of commentary: preparatory elements, perceptual elements, and productive elements. For each category, we summarize the recommendations from the e-list participants.

Preparatory Elements

1. Focus on breathing and breath control, and on explosion rather than implosion. Time spent on the airstream mechanism is particularly valuable, as the implosive nature of Vietnamese is in direct contrast to the explosive nature of English.
2. Recognize and address the glottal stop, a common phonemic feature of Vietnamese, which is embedded in two of its six tones. The glottal stop interferes with English pronunciation, particularly in the enunciation of syllable-final consonants as well as with the connected speech features of English. This tendency for glottal stop insertion in English distracts listeners from the message. Understanding and gaining awareness of the occurrence of glottal stops is fundamental to helping learners avoid them.

Perceptual Elements

3. Focus on auditory perception before oral production. Discussion ensued about how to get students to *hear* the correct pronunciation, no matter whether segmental sounds or pitch patterns. Strategies include telling listeners what to listen for, modeling the target pronunciation feature, getting learners' own speaking output to converge on the target, and forming a closed-circuit auditory feedback loop.
4. Lead students to *hear* the "correct" pronunciation by *producing* the target pronunciation through a variety of means, such as singing, sagittal images of the articulators, verbal description, modeling.
5. Encourage students to hear English without looking at the written text to focus their attention on the actual sounds of the language. Sound-symbol correspondence is extremely strong in Vietnamese and therefore helpful, but considerably weaker in English and often misleading. Considering English an *ear* language as opposed to Vietnamese as an *eye* language may aid learners in framing their listening.
6. Have students sing as a means of aiding auditory perception. (See expansion below.)

17

Productive Elements

7. Tackle consonants, particularly finals (e.g., /l/-/w/ *feel–few*, /l/-/n/ *fall–fawn*, /t/-/s/ *pat–pass*) along with the deletion of consonants and insertion of glottal stops, as in /kɔʔ ən ɪfɛʔ/ for "cause and effect". The omission or inaccurate articulation of consonant clusters–which do not exist in spoken Vietnamese–is particularly problematic in English and worthy of attention.
8. Focus on English word stress, phrase stress, and pitch patterns.
9. Avoid techniques that may backfire with learners. For example, telling a student to simply "Repeat after me" or "Say it the way I say it" may result in the student replying, "But I *am* saying it that way," illustrating an inability on the part of the student to perceive the intended target.
10. Refrain from pointing out the specific shortcomings of a learner's speech, as this may lead to hyper-correction; instead focus on the learners' hearing the "correct" pronunciation.[7]

The Use of Singing to Improve Learners' Pronunciation

Overlapping the three general categories and woven throughout this thread was a discussion of the efficacy of singing to improve pronunciation.[8] Songs can enable learners to gain a feel for English, create a motor memory, and enhance prosodic elements of spoken language. The discussant with a large proportion of Vietnamese learners of English shared a link to student recordings of a song used in teaching stress, rhythm, linking, and selected segmentals.[9] Along with an explanation of the task, she also provided a brief commentary on 10 Vietnamese students' renditions of the song, making observations on their insertion of glottal stops and nasals, their overall stress, rhythm, intonation, articulation of consonants, and linking to connect words in phrases. The initiator of this thread recounted how the intelligibility of one Vietnamese speaker improved after she had him sing part of a song and then speak the lyrics, focusing on the airstream mechanism. She closed with an expression of appreciation for the helpful discussion on her initial posting.

Topic 2: Stress Shifting in British and American English

British English (BrE) and American English (AmE) often stress the same printed word differently. Are there any patterns or trends? This thread began when a North American discussant who had been watching a historical documentary on the ancient Germanic tribes observed that the British female narrator of the series repeatedly placed stress on the second

---

[7] As previously mentioned, relying on the written form was noted as a deterrent, as English orthography can confuse rather than promote proper pronunciation.

[8] Circling back to an earlier comment that native speakers of a tone language have a higher incidence of perfect/relative pitch–from the perfect vs. relative pitch "hot topic" from last year's study (Chan & Brinton 2015)–the teachers of English to Vietnamese speakers in this group stated that, in their experience, speakers of Vietnamese, a tone language, do not have better L2 prosody or better comprehensibility than speakers of non-tonal languages.

[9] See http://www.voxopop.com/topic/34a6866f-7d31-4126-a114-07e2117f1f19.

syllable of the word *priestess*: **pries`TESS**, yet she did not stress all words with this suffix on the final syllable; curious, he posted this for discussion.

Priestess-Princess

1. ***Longman Pronunciation Dictionary*** (Wells, 2008) gives stress on the second syllable of *priestess* as the preferred pronunciation for BrE (but not for AmE). Both this source and the ***English Pronouncing Dictionary*** (Jones, 2011) show stress on the suffix as the preferred pronunciation in BrE, but not for AmE.[10]
2. Younger people have a greater tendency than older people to stress the first syllable, suggesting that this is an ongoing sound change; still, 60% in a 1998 poll preferred **prin`CESS** over **`PRIN-cess**. Thus, the argument continued, if the trends of *princess* and *priestess* are similar, the first syllable of both will eventually predominate in Britain, representing a generational shift. (BrE practitioner in Southeast Asia)
3. The stress on *princess* depends on context: She's a **prin`CESS**, but when followed by a stressed word, such as a name, the stress shifts to the first syllable: **`PRIN-cess Di-AN-a**. (BrE practitioner in Britain)

French-English influences

4. Britons tend to stress words that look French on the last syllable while AmE tends to shift the stress forward. Examples (BrE/AmE): **em-plo`YEE/em`PLOY-ee, lem-on`ADE/ `LEM-on-ade, prin`CESS/ `PRIN-cess. (**BrE practitioner in France)
5. BrE speakers may change more French words to conform to English stress than AmE speakers do. Examples: **`GARage/ga`RAGE, `MASsage/mas`SAGE.** (AmE practitioner in the US)
6. Some borrowed vocabulary items from French are given different stress patterns in BrE and AmE. This may reveal more about these two varieties of English than it does about French, which gives more or less equal stress to each syllable. (BrE practitioner in Britain)
7. As French words become Anglicized, the stress moves to the first syllable.
    Example Set A: **`VIL-lage, `MARriage, `CAR-riage,** and **`VOY-age** all now have initial stress (except in **Bon vo`YAGE**, a set phrase).
    Example Set B: **col`LAGE** and **mas`SAGE** still have final stress.
    Example Set C: the stress patterns in **gar`AGE/ `GARage** are variable.
    One would predict that, over time, all of these nouns will have initial stress. An additional factor and exception to this prediction: a long vowel or otherwise "heavy" syllable may prevail.
    Example Set D: **mas`SAGE, u`NIQUE**, which have a long vowel in the second syllable, and **pictur`ESQUE**, which ends in /sk/, a consonant cluster, retain final syllable stress. (BrE practitioner in Southeast Asia)

---

[10]*Dictionary.com* (©2016 Dictionary.com LLC) and *Merriam-Webster.com* (©2015 Merriam-Webster, Inc.) give only one pronunciation with the stress on the root: **`PRIES-tess**.

Noun/Verb distinctions

A third category of stress shifts discussed in this thread revolves around nouns and verbs that have the same spelling, e.g., *permit*.

8. *Permit*: The LPD shows BrE **PER-mit** for the noun and **per`MIT** for the verb; the same is true for AmE, but it also shows that for some Americans **per`MIT** is also used for the noun.
9. Language change has occurred over the centuries, and *permit* is following a trend affecting a wide range of other words. [11] Aitchison (2013) states that in 1570 a dictionary listed just three words for which the stress on the noun had shifted to the first syllable, while the verb retained stress on the second syllable: **`REB-el/re`BEL**, **`REC-ord/re`CORD**, and **`OUT-law/out`LAW**. By 1582, a further five items were added, and by 1932, the number was 150. Aitchison gives *recess* as an example of a noun that has recently undergone this stress shift for nearly everyone in America but not everyone in Britain.
10. *Address*: BrE does not generally have a noun/verb distinction; both are pronounced **ad`DRESS**. On the other hand, AmE prefers **`AD-dress** for the noun and **ad`DRESS** for the verb.
11. *Research* is another word undergoing a change. AmE and Estuary English (a variety of English widely spoken in Southeast England), both tend to follow this noun/verb distinction: **`RE-search/re`SEARCH**. Estuary English is more porous of AmE pronunciation influences than the more standard type of BrE speech.

Concluding this thread, several participants predicted that with AmE leading the way, other varieties will follow this trend of shifting stress from the final to the initial syllable to create a distinction between noun and verb pairs.

Topic 3: Vowel Charts

This exchange originated with a query from a pronunciation specialist proposing and seeking feedback on a new vowel sound chart. Still in its prototype version, as described by its originator the chart sought to "represent vowel length in concentric rings and vowel quality in radiating spokes" (see Figure 1).

Reactions and suggestions to the chart varied, with some participants providing concrete feedback, some suggesting references to consult, others providing a global assessment (either positive or negative), and yet others questioning the rationale behind the new chart:

1. "My intro linguistics students would have found it wildly confusing."

---

[11] Stockwell and Minkova (2001) refer to this as a systemic exception to the general tendency in English for two-syllable words to be stressed on the first syllable. They note over 130 pairs of words in English which have different stress patterns depending on whether they are functioning as nouns or verbs (with additional instances of adjectives which have the same stress pattern as the corresponding noun). Compare: pre`SENT (verb), `PREsent (noun), `PREsent (adjective).

2. Revise the chart in accordance with the 8th edition of *Gimson's Pronunciation of English* (Cruttenden, 2014) to eliminate vowels (e.g., /æ/) which are disappearing from General British. Also, it would need to be radically revised for American English.
3. What is gained by moving away from the conventional mouth formation related (MFR) chart (see Figure 1), which signals the general locations in the mouth where sound formation begins?
4. The assignment of color in the chart appears random. Why use color? It is not useful for those who are color blind and potentially confusing for those who have synesthesia.[12]
5. Why is /ɛ/ not used to represent the vowel in *leg*? The vowel at 8 o'clock should read *eye*, not *ye*. The diphthong in *tourist* is missing. Also absent from the chart is the movement associated with diphthongs.
6. It's confusing to have so many vowels represented with /r/. Why not use postvocalic /n/ or some other coda?

In response to the suggestions, the originator of the chart responded as follows:

1. To avoid confusion, students could be introduced to the chart on a need-to-know basis.
2. The chart is meant to be compatible with British learner dictionaries; hence /æ/ is retained and /e/ is used in place of /ɛ/. Admittedly, it would require substantial revision of the chart for it to represent American English.
3. The MFR chart does *not* capture the six long/short vowel pairs. The proposed chart *does*; in addition to vowel length, it captures the idea of vowel quality.
4. Currently, I am just "experimenting" with the color aspects of the chart. It could potentially be used to reinforce the patterning of the chart.
5. Postvocalic /r/ is crucial in British English long vowels such as *shirt* and *hair*; other codas are not possible.
6. The diphthong in *tourist* was intentionally omitted as it has coalesced in General British with the vowel in *fork*.
7. The proposed vowel sound chart allows for regional variability. This is an advantage over the MFR chart, which situates the vowel in a fixed location and thereby implies precision of articulation.

---
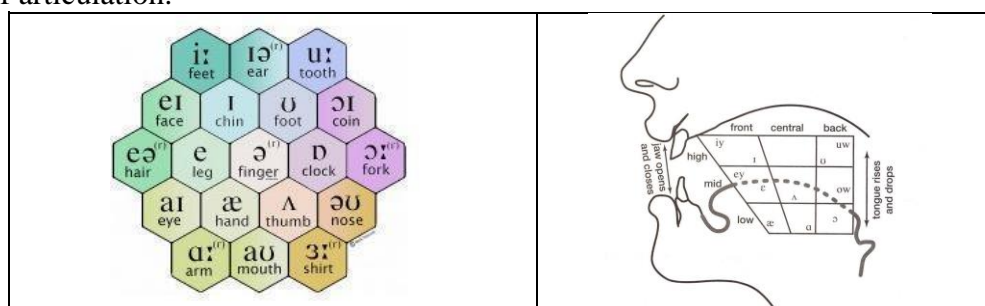
[12] In fact, a side discussion about the condition of *synesthesia* (the involuntary co-occurrence of a second sensory perception in conjunction with a first; for example, hearing a sound and visualizing an associated color) occurred as a result of one of the discussants describing his personal experiences with this condition. We've opted not to synthesize that portion of the discussion, in which several participants cited references on synesthesia.

Figure 1. A comparison of the proposed vowel sound chart and the MFR vowel chart

In addition to the suggestions and responses summarized above, the strand contained additional discussion of Gattegno's original sound color chart (1985) and recently revised versions of his chart (Teaching the Silent Way, n.d.), Underhill's (n.d.) interactive phonemic chart, and Taylor and Thompson's color vowel chart (English Language Training Solutions, n.d.), with various discussants extolling the merits or describing the challenges of using each.

Topic 4: The Contrastive Analysis Hypothesis

The e-list posting which initiated this discussion strand posed the question of whether the contrastive analysis hypothesis (CAH) (Lado, 1957)[13] has current value as a research framework or if, instead, it "may be somewhat helpful for a teacher but is simply too inexact for research." By way of general comments, discussants noted that it is important to keep in mind the distinction between the "weak" and "strong" versions of CAH (Wardaugh, 1970)[14] and whether the model is being applied to production or perception (the original model was intended to describe perceptual processes).

Comments arguing for the value of the CAH in pronunciation teaching and research included the following:

1. Contrastive analysis is needed and valuable, if underestimated.
2. It may not explain all errors (e.g., those due to the developmental processes of interlanguage), but to a large extent it reliably predicts features of a foreign accent.
3. Flege's speech learning model (SLM) (Flege, 1995)[15] is more credible than the CAH because it adds a third (i.e., "similar") category to the binary "same" vs. "different" categories of the CAH; however, it relies on experimental verification of the same/similar/different categories and is therefore more demanding than CA.
4. CA is a useful way of looking at the issue of transfer. There are many useful works written that apply CA cross linguistically, including work on prosody (e.g., Hirst & Di Cristo, 1998).

---

[13] The CAH claims that the principal barrier to second language acquisition is interference from the first language system with the second language system (Lado, 1957). As an example, German final stop consonants are devoiced; hence the CAH would claim that German speakers learning English would have difficulty voicing final stop consonants.

[14] In his influential 1970 article, Wardaugh distinguishes between a "strong" version of CAH, applied to predict areas of learner difficulty in the L2 and a "weak" version wherein CA is seen to have explanatory power (i.e., it can be applied after the fact to help explain learner difficulties) but lacks predictive power.

[15] Flege's SLM proposes that the more distant a second language sound (phonetic segment) is from the closest first language speech sound, the more learnable the second language sound will be.

In addition to Flege's SLM, several other hypotheses were proposed as alternatives to the CAH. These included: Eckman's (1977) markedness differential hypothesis;[16] Best's (1994) perceptual assimilation model; Kuhl & Iverson's (1995) perceptual magnet effect; and MacWhinney's (2008) unified model.[17] All of the above models (including the SLM), it was noted, require a satisfactory operationalization of what it means for sounds in two languages to be the same, similar, or different.

The consensus of this discussion strand is best summarized by the following comment from an e-list participant who had not been otherwise active in this discussion strand: "There was a time when CAH was considered 'politically incorrect,' but a contrastive study of the phonologies of different languages yields very valuable information for the linguist and the teacher. I was not particularly interested in absolute accuracy of FL pronunciation, but in intelligibility, and what interferes with intelligible pronunciation/perception. CA (if not CAH) provides invaluable information."

DISCUSSION AND CONCLUSIONS

In determining the "hot" topics discussed in this article, we have taken into account not only the quantity of responses (i.e., number of discussants and exchanges) but also the quality (i.e., general or global interest and depth of discussion). Thus, while not all the most voluminous in the discussion thread, the four topics we chose as having the greatest global interest, depth of discussion, and/or relationship to teaching pronunciation in this one-year period are Vietnamese speakers, British English and American English stress patterns, vowel charts for learners, and contrastive analysis.

E-list participation by pronunciation specialists can yield edifying discussions. For one, such a discussion forum allows practitioners to post queries and receive feedback from a worldwide network of members with a variety of teaching environments and experience researching different aspects of pronunciation. The forum also allows for open discussion among professionals, providing a relatively safe place to propose a new or different concept or interpretation and to critique each other's ideas. In addition, it brings heretofore unheard of ideas to some, while confirming and reassuring familiar ideas to others. As in our previous work, this article demonstrates the value of communication amongst practitioners on the chosen e-list.

---

[16] The markedness differential hypothesis represents Eckman's revision of the CAH. He proposes that those aspects of the second language that are more marked (less dominant or regular) than the first language will be the most difficult for learners to acquire (i.e., with the degree of difficulty depending on the degree of markedness). On the other hand, those aspects which are different but not marked will be easier to acquire. As an example, the syllable structure consonant + vowel (CV) is the most common (or unmarked) universally. Adding a consonant (e.g., CVC) or consonants (e.g., CVCCC) adds a degree of markedness and hence difficulty to the syllable structure for learners from a CV language. Hence English *bee* (CV) is unmarked; however, the plural form *bees* (CVC) is marked. In increasing degrees of markedness (and therefore difficulty) we would predict that the more marked forms *beast* (CVCC) and *beasts* (CVCCC) would be even more difficult for a learner from a CV language to pronounce.

[17] For an explanation of the perceptual assimilation model, the perceptual magnet effect; and the unified model, readers are referred to the original sources as an explanation of these theories is beyond the scope of this paper.

ACKNOWLEDGMENTS

ABOUT THE AUTHORS

Marsha J. Chan is an educational consultant, 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 200 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.

Donna M. Brinton is an educational consultant based in Beverly Hills, California. She has taught on the TESOL/Applied Linguistics faculties of the University of Southern California, Soka University of America, and the University of California, Los Angeles, where she also served as the coordinator of the university's English as a Second Language program. She has written and co-edited numerous professional texts and is one of the authors of *Teaching Pronunciation: A Course Book and Reference Guide* (Cambridge University Press, 2010). Donna frequently presents on the topic of practical phonetics at national and international conferences. Her interest lies in the practical application of phonetics to the second/foreign language classroom and in helping prepare teachers to teach pronunciation.

REFERENCES

Aitchison, J. (2013). *Language change: Progress or decay?* (4th ed). New York, NY: Cambridge University Press.

Best, C. T. (1994). The emergence of native language phonological influences in infants: A perceptual assimilation model. In J. C. Goodman and H. C. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 167-224). Boston, MA: MIT Press.

Brinton, D. M., & Goodwin, J. (with Celce-Murcia, M). (2006, December). World English, intelligibility, and pronunciation standards: What pronunciation specialists think. *Speak Out!*, *36*, 26-32.

Brinton, D., & Chan, M. (2015). What's hot, what's not: Insights form pronunciation practitioners. In J. Levis, R. Mohamed, M. Qian & Z. Zhou (Eds), *Proceedings of the 6th Pronunciation in Second Language Learning and Teaching Conference* (pp. 154-165). Ames, IA: Iowa State University.

Cruttenden, A. (2014). *Gimson's pronunciation of English* (8th ed.). New York, NY: Routledge.

Gattegno, C. (1985). *The learning and teaching of foreign languages*. New York: Educational Solutions.

Jones, D. (Ed.). (2011). *English pronouncing dictionary* (15th ed.). Cambridge, England: Cambridge University Press.

Eckman, F. (1977). Markedness and the contrastive analysis hypothesis. *Language Learning*, *27*(2), 315-330.

English Language Training Solutions. (n.d.). *Color vowel chart*. Retrieved from http://elts.solutions/color-vowel-chart/

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.

Hirst, D., & Di Cristo, A. (1998). (Eds.). *Intonation systems: A survey of twenty languages*. Cambridge, England: Cambridge University Press.

Kuhl, P. K., & Iverson, P. (1995). Linguistic experience and the "perceptual magnet effect." In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 121-154). Timonium, MD: York Press.

Lado, R. (1957). *Linguistics across cultures*. Ann Arbor, MI: University of Michigan Press.

MacWhinney, B. (2008). A unified model. In P. Robinson & N. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 341-371). New York, NY: Routledge.

priestess. (n.d.). *Dictionary.com Unabridged.* Retrieved February 19, 2016 from Dictionary.com website http://dictionary.reference.com/browse/priestess

priestess. (n.d.). *Merriam-Webster.* Retrieved February 19, 2016 from Merriam-Webster.com website http://www.merriam-webster.com/dictionary/priestess

Stockwell, R., & Minkova, D. (2001). English words: History and structure. New York, NY: Cambridge University Press.

Teaching the Silent Way. (n.d.). Retrieved from http://www.teachsw.org/the-sound-colour-chart

Underhill, A. (n.d.) *Interactive phonemic chart*. Retrieved from http://www.onestopenglish.com/skills/pronunciation/phonemic-chart-and-app/interactive-phonemic-chart-british-english/

Wardaugh, R. (1970). The contrastive analysis hypothesis. *TESOL Quarterly*, *4*, 123-130.

Wells, J. C. (Ed.). (2008). *Longman pronunciation dictionary* (3rd ed.). Harlow, England: Pearson Education.

APPENDIX

Sources Provided by E-list Participants

Vietnamese Speakers

Chan, M. J. (2010, June 3). Voxopop for Speaking Pronunciation and Listening [Webinar]. Retrieved from https://sas.elluminate.com/site/external/jwsdetect/playback.jnlp?sid=&psid=2010-06-03.0958.M.953CC911BA4E76CCBF1E1A8C81AAE4.vcr

Chan, M. J. (2010, October 21). Using Voxopop for speaking, pronunciation and listening [Webinar]. Retrieved from http://www.sunburstmedia.com/present/Chan_Voxopop_webinar.pdf

Chan, M. J. (2011, October 21). *My Bonnie lies over the ocean.* [Video file]. Retrieved from http://www.voxopop.com/topic/34a6866f-7d31-4126-a114-07e2117f1f19#.Vh05a9Z-GIE

Gilbert, J. B. (2012). *Clear speech* (3rd ed.). New York, NY: Cambridge University Press.

Media Mill. (2010, October 21). *Lam*. [Video file]. University of Minnesota: Minneapolis, MN. Retrieved from http://mediamill.cla.umn.edu/mediamill/embed.php?subclip_id=3815&live=true&chrome=true

Stress Shifting in British and American English

Aitchison, J. (2013). *Language change: Progress or decay?* (4th ed). New York, NY: Cambridge University Press.

Harry Reser's Syncopators. (2009, July 23). *I scream - you scream - we all scream for ice cream.* [Audio file]. Retrieved from https://www.youtube.com/watch?v=o1NXup3hDm8

Jarmusch, J. (2007, May 28). *"Down by law." Episode: I scream, you scream, we all scream for ice cream*. [Video file]. Retrieved from https://www.youtube.com/watch?v=7rK3s_BP9kE

Jones, D. (Ed.). (2011). *English pronouncing dictionary* (15th ed.). Cambridge, England: Cambridge University Press.

priestess. (n.d.). *Dictionary.com Unabridged.* Retrieved February 19, 2016 from Dictionary.com website http://dictionary.reference.com/browse/priestess

priestess. (n.d.). *Merriam-Webster.* Retrieved February 19, 2016 from Merriam-Webster.com website http://dictionary.reference.com/browse/priestess

Wells, J. C. (Ed.). (2008). *Longman pronunciation dictionary* (3rd ed.). Harlow, England: Pearson Education.

Vowel Charts

Brown, A. (2014). Pronunciation and phonetics: A practical guide for English language teachers. New York: Routledge.

Cruttenden, A. (2014). *Gimson's pronunciation of English* (8th ed.). New York, NY: Routledge.

Cytowic, R. E. (1998). *The man who tasted shapes*. Cambridge, MA: MIT Press.

English Language Training Solutions. (n.d.). *Color vowel chart*. Retrieved from http://elts.solutions/color-vowel-chart/

Frost, D. (n.d.). *Kinephones*. Retrieved from http://kinephones.u-grenoble3.fr/#/en_gb/1/table

Gattegno, C. (1985). *The learning and teaching of foreign languages*. New York: Educational Solutions.

Hancock, M. (2015, January 7). *Mark's vowel chart*. [Blog post]. Retrieved from http://hancockmcdonald.com/blog/marks-vowel-sound-chart

Haptic-integrated clinical pronunciation research and teaching. (2011, November 23). *The color and felt sense of English vowels*. [Blog post}. Retrieved from http://hipoeces.blogspot.ca/2011/11/color-and-felt-sense-of-english-vowels.html

Mompeán Guillamón, P. (2013). Vowel-color symbolism in English and Arabic: A comparative study. *Miscelánea*, *47*, 31-52.

Sachs, O. (2008). Musicophilia: Tales of music and the brain. New York: Vintage.

Teaching the Silent Way. (n.d.). Retrieved from http://www.teachsw.org/the-sound-colour-chart

Underhill, A. (n.d.) *Interactive phonemic chart*. Retrieved from http://www.onestopenglish.com/skills/pronunciation/phonemic-chart-and-app/interactive-phonemic-chart-british-english/

Contrastive Analysis

Best, C. T. (1994). The emergence of native language phonological influences in infants: A perceptual assimilation model. In J. C. Goodman and H. C. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 167-224). Boston, MA: MIT Press.

Corder, S. P. (1975). Error analysis, interlanguage and second language acquisition. *Language Teaching*, *8*(4), 201-218.

Delattre, P. (1965). Comparing the phonetic features of English, French, German and Spanish. Heidelberg, Germany: Groos.

Eckman, F. (1977). Markedness and the contrastive analysis hypothesis. *Language Learning*, *27*(2), 315-330.

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.

Hirst, D., & Di Cristo, A. (1998). (Eds.). *Intonation systems: A survey of twenty languages*. Cambridge, England: Cambridge University Press.

Kuhl, P. K., & Iverson, P. (1995). Linguistic experience and the "perceptual magnet effect." In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 121-154). Timonium, MD: York Press.

Lado, R. (1957). *Linguistics across cultures*. Ann Arbor, MI: University of Michigan Press.

MacWhinney, B. (2008). A unified model. In P. Robinson & N. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 341-371). New York, NY: Routledge.

Peperkamp, S., & Dupoux, E. (2002). A typological study of stress "deafness." In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology 7* (pp. 203-240). Berlin, Germany: Mouton de Gruyter.

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

Swan, M., & Smith, B. (Eds.). (2001). *Learner English: A teacher's guide to interference and other problems* (2nd ed.). Cambridge, England: Cambridge University Press.

Thomson, R. I., Nearey, T. M., & Derwing, T. M. (2009). A modified statistical pattern recognition approach to measuring the crosslinguistic similarity of Mandarin and English vowels. *Journal of the Acoustical Society of America, 126,* 1447-1460.

Trubetzkoy, N. S. (1939). Grundzüge der Phonologie. Prague, Czech Republic: Cercle Linguistique de Prague.

Wardaugh, R. (1970). The contrastive analysis hypothesis. *TESOL Quarterly*, *4*, 123-130.

IN OTHER PEOPLE'S WORDS: LANGUAGE LEARNERS' IMITATION OF PROFESSIONAL SPEECH

Janay Crabtree, University of Virginia

This paper explores international graduate student language learners' (LLs) (N=17) enjoyment and perceptions of improvement when using voiceovers of TED Talks as a pronunciation-improvement tool in an academic communications course through a survey (see *Appendix A*). One major finding is that the majority of the LLs contrasted themselves to native speakers and make statements of dissatisfaction with their oral proficiency or pronunciation even though the majority of LLs enjoyed the activity and felt they had improved their pronunciation. A discussion of how these findings of dissatisfaction fit into LLs "noticing the gap" (Schmidt, 1990) or raising awareness of their pronunciation needs follows. Implications for LLs' identities are also explored.

INTRODUCTION

Voiceovers have been discussed as an effective and enjoyable teaching tool for language learners (Henrichsen, PSLLT 2014), and studies in shadowing (repetition of listening to model speech) also suggest imitative speech is an effective comprehension and retention tool in L2 learning (de Guerrero and Commander, 2013). International teaching assistants (ITAs) are often asked to record and imitate the speech of native-English speakers as a pronunciation exercise (Goodwin, 2007), and imitation and repetition are often proposed in teacher training materials (Harmer, 2012). However, little research has been conducted on how language learners (LLs) take up imitation activities such as these and internalize them in terms of their pronunciation development and identity. The purpose of this study, which consisted of a survey and an interview of seventeen students who participated in a TED Talk (TT) voiceover exercise in two academic communications courses, stems from the practice of advising language students to imitate native speakers of the target language to become more orally proficient speakers, particularly in terms of L2 pronunciation. This study asks the following questions:

1. How do LLs perceive imitation activities, in particular the specific TT voiceover imitation activity?

2. Do LLs enjoy the voiceover activity, in particular do they perceive this activity as valuable, specifically in improving pronunciation skills?

3. If LLs perceive self-improvement, how do LLs utilize the learning and practice from this exercise to continue pronunciation improvement—if they do?

Participants

The participants in this study were adult learners in a support program for graduate students who have been accepted and matriculated into a Southeastern U.S. university. The learners are filtered into oral classes through one of the two different assessments: an in-house interview or the Speaking Proficiency English Assessment Kit (SPEAK) for those planning to be International Teaching Assistants. The majority of the LLs were in science, technology, engineering, math (STEM) or in accounting/commerce. In many cases, the students opt to take the course, particularly in the summer. In some cases, however, the students are recommended, and in some cases required by their departments, to continue language support classes to work on proficiency for one of the following reasons: 1. Successful participation in content classrooms, 2. Continued work toward the minimum score of 55 on the SPEAK test or completion of an ESL advanced academic communication course to be an ITA (the 5th and final level of oral communications class offered in the ESL program), 3. Completion of a personal goal, such as presenting at professional conferences. On average, the LLs are over twenty two years of age and have studied English for over ten years, are advanced learners of English, and have scores well above the minimum of 90 composite on the ibt TOEFL which is the recommended minimum for acceptance into most university departments. Almost all of the students in these courses test above the recommended score for beginning level of oral academic communication.

METHODS

Henrichsen (PSLLT, 2014) reported the practice of video voiceovers "to extend the benefits of researched pronunciation practice and tracking" and found that his students enjoyed voicing over animated films. Initially, this research was developed to pilot the introduction of the voice-over activity and help instructors determine whether they desired to continue the voice-over activity as part of the oral academic course curriculum. In order to answer the above research questions concerning enjoyment of the voiceover exercise as well as if the participants perceive it as an activity that facilitates pronunciation improvement, graduate students in two ESL oral academic communications courses were recruited to be participants in this study.

The oral academic communications classes are skills classes; however, within this class, students focus on pronunciation and vocabulary structures as needed. In doing so, the students are asked to discuss what they think makes *comprehensible* and *intelligible* speech and communication and thusly, *comprehensible* and *intelligible* pronunciation, prior to their first presentations. The terms *comprehensible* and *intelligible* (Munro & Derwing, 1995) are specifically used with the students in the first two weeks of class, in the assignment sheet, as well as on the feedback sheets they receive from their introductory presentations. In defining comprehensibility with the students, the instructors introduce elements and examples found to be related to comprehensibility such as effective thought groups including focus within them, linking, word stress, segments, intonation and rhythm, speed, and contrastive stress. In the first presentation, the students are asked to bring a transcript of their introductions to specifically discuss their comprehensibility and the terms that comprise it as discussed in class, and this is noted in both the activity's description on the syllabus, the assignment description listed in the collaborative site, and on the rubric or feedback sheet. All of these terms are defined for the participants and later the participants are asked to apply these definitions to the final presentation, particularly in marking transcripts for effective thought groups, focus, stress—both within words and contrastively in the discourse,

linking and rhythm as well as reductions, segments they have difficulty with and want to practice, and intonation.

While the research that the terms "comprehensibility" and "intelligibility" originate (Munro & Derwing, 1995) is not shared with the LLs, these terms are used and discussed in the following ways, when the LLs review their transcribed introductions with partners (language consultants who volunteer in the ESL classes): "How well do your group members understand you? How hard do your group members have to work in order to understand you?" "Are you monitoring your group members' faces for understanding?" "What body language tells you your group members are comprehending you?" The students also often note that comprehension is determined by many factors that are non-linguistic, such as eye contact and body language. These elements are also discussed as affecting communication competence in the course and especially in presentations.

The academic communications course implemented a TED-talk voiceover as part of the curriculum in 2014-15. In two academic oral ESL courses, graduate student LLs were asked to take part in short (less than 3 minutes) TT "voiceovers" as part of the curriculum of these pass or fail classes. The language learners (LLs) self-selected a talk and focused on practicing their TT during the last ten weeks of the semester. All of the students, with the exception of one, chose TTs with Native-English speakers as their models. Each student sent the instructor of the course an initial MP3 of their reading of the TT between the second and third week of the course, approximately ten weeks prior to presentation, so that the instructor could compare the initial recording to the final presentation which occurred after LLs had practiced inside and outside of the course. Students were surveyed (see *Appendix A*), and follow-up interviews were conducted with those students who were successfully recruited for the research. Since the students in this program are graduate students matriculated into departments, and many of these students have the expectation of eventually serving as ITAs, professional speeches were utilized as the voice-over material. Since TTs have a variety of speakers, topics, transcripts, and closed-caption ability, but are often performed by professionals, particularly academics who are professionals in their field, the TTs were deemed an appropriate fit for voice-over materials for this particular population, a population with a majority that will most likely find itself giving academic lectures of some sort. Language consultants (native-speaking volunteers who work in the oral classes) often participate in peer or pair work within classes from week two to twelve. The LLs worked with language consultants the last few minutes of every class to discuss unknown words, stress patterns, meanings, stress, focus, thought groups, linking, and intonation of certain discourse units. LLs were asked to bring their transcripts of their chosen TT to their oral academic course to practice if time permitted at the end of every class. The LLs were also recommended to practice outside of class and were taught about shadowing. LLs participated in class discussions of different methods for shadowing with various recording tools such as audacity, multi-track song recorder (MTSR), or merely using the TED Talk and a hand-held recorder. The LLs were asked to practice shadowing, but shadowing was not mandated as the sole method of practice, as I was interested in how the LLs would choose to practice or review the TT.

The LLs were video recorded while performing their voiceovers in the last week of classes. All students were asked to view their presentations which were made available through the course site. LLs were asked to then fill out a survey and take part in an exit interview concerning the total voiceover activity.

Seventeen of the eighteen student LLs participated in the activity, filled out surveys, including a scalar item concerning self-perception of enjoyment and improvement of the activity. Finally, LLs underwent exit interviews that were a follow-up to the information they provided in the surveys about the activity (see *Appendix A*). Interviews and surveys were then partially transcribed and coded for possible shared themes or phenomena (Crotty, 2003) in which the LLs constructed how they understood and enjoyed the activity. They were also asked specifically about how they perceived the activity's results, particularly in terms of effect for pronunciation.

RESULTS

The following results from this research will be discussed: 1. The overall perception of enjoyment versus improvement from all LLs; 2. Qualitative data from surveys and interviews that suggest the LLs compare themselves to the TED speakers; and 3. Statements in the qualitative data from the LLs that suggest some of them did not understand elements making up the term *pronunciation.*

The majority of learners enjoyed the activity; only one LL rated the activity a 1—the lowest ranking for enjoyment of this activity. Eleven LLs enjoyed the activity even more than they thought they had improved their pronunciation, and four of the LLs ranked enjoyment and improvement equally. The learners all perceived they had improved, even if incrementally, even the LL (KNW in *Figure 1*) who rated the least enjoyment or a 1 on the rating scale of 1-5 rated his improvement as a 3 out of the rating scale 1-5, which suggests that while he did not necessarily enjoy the activity, he perceived that it helped him improve his pronunciation (see *Figure 1*).
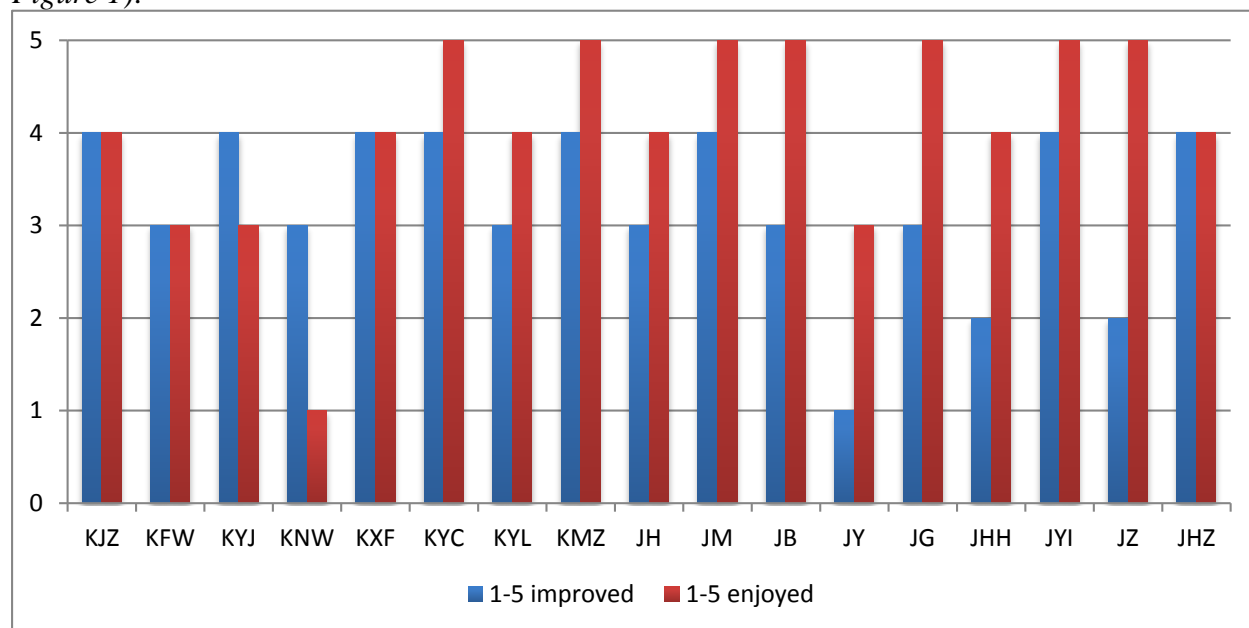

Figure 1. Participants perceived improvement vs. enjoyment.

Although participants' marking of the scalar items suggests they perceived they had improved their pronunciation, they also made statements that reflected conflicting feelings revealed in the discourse when they were asked to review their presentations. Especially noteworthy is that they

did not concentrate solely on their own performance, but their performance in comparison to the professional speaker. The LLs' discourse is full of contradictory statements in which the LLs feel they have improved on the one hand, but are still dissatisfied with their proficiency on the other hand. For example, one student wrote in the survey "though I felt helped by practicing the speech, I realized it is still a long way to go to be a proficient English speaker to address a speech precisely" (KYL, 2015 see *Appendix B*) Another LL reflected the following: "My voice is dim, and I used to think that [when] I would do the presentation, I would try to sound really clear, after, I see the videos, I feel a little disappointed [in] myself (JB, 2015 see *Appendix B*)." The previous types of discourse display that the LLs may be comparing themselves to the professional speakers and perceiving themselves as less proficient through that comparison. In other words, much of the LLs' discourse does not celebrate the perceived gains they have made from their first recording to their final presentation, as noted in the survey results and presented in table 1, but, instead, the discourse focuses on what they perceive defective in comparison to the native-speaking professional speakers.

A final result of this research is that not all of the LLs may understand the term *pronunciation* and what it entails even though this term is used in the course and course objectives. In discussing this term, *comprehensibility* and *intelligibility* are targeted as goals for the course. Even though the students discuss and practice elements that comprise pronunciation such as thought groups, intonation (rising, falling, and level), stress within words, contrastive stress, focus within thought groups, rhythm, linking, reduction, speed, and segments (called individual sounds in the course), they still may view their pronunciation as limited to individual sounds as evidenced by LLs' statements in the interview or survey. For example, LLs stated they perceived they had made improvements in "intonation," "stress in words," "pitch," "volume," "rhythm" and "speed," but that they did not feel they had improved *pronunciation*, even though the majority of the students (over 80%) marked they had improved their pronunciation 3-4 on the scale. No one marked the scale as a 5 (the highest rating) in terms of their improvement on pronunciation. The discourse concerning pronunciation of the LLs needs to be further unpacked, as it is unclear whether LLs do not understand that these elements (rhythm, intonation, stress, etc.…) are part and parcel of pronunciation or if the LLs perceive they are improving in specific elements of pronunciation but not overall or holistically. Contrastively, LLs might understand these terms as comprising pronunciation, but they may have felt that they did not improve enough in all of these elements, even when they marked themselves as improving their pronunciation on the survey scale.

CONCLUSION

The following items become evident from this research. First, students may not understand what makes up the term *pronunciation,* even though the terms *comprehensible* and *intelligible* were utilized in the discussion of pronunciation prior to their first presentation, and terms such as thought groups, focal stress in thought groups, stress within words, contrastive stress, individual sounds (segments), intonation, linking, speed, rhythm, and reduction were introduced and practiced in relation to the TTs. The topics in the course, such as use of stress in key words and contrastive stress or understanding of linking in discourse groups to help listeners chunk the information are utilized to improve comprehensibility of the LLs, but it is unclear if the LLs understand these are connected to pronunciation as well, or if the LL still merely perceives

pronunciation as individual sounds or segments. A future survey will need to be developed to understand the LLs' perceptions of this term better.

Secondly, teachers of oral English may need to consider using more non-native speaking models in order to motivate LLs and provide guides for effective comprehensibility, not perfect pronunciation. The LLs self-selected their TTs, and many stated they selected for field, topic, or length, but utilizing non-native models for discussion prior to the activity might serve as "aspirational models" (Murphy, 2014, p. 259) for LLs. Non-native speaking models may assist LLs in further understanding comprehensible pronunciation and also help them to not have unrealistic expectations. However, this recommendation may be connected to LLs' interpretation of the term *pronunciation* and *comprehensibility*. LLs may not understand the term *pronunciation,* and the term may need to be unpacked even more, particularly as many LLs stated they improved intonation, speed, rhythm, and stress, but not pronunciation. However, the misunderstanding of the term also may be an effect of practicing various features of pronunciation on different days (i.e. rhythm for thought groups during week four, contrastive stress on week five, etc.). After all, pronunciation is a complex term with many features comprising it. While a majority of the LLs enjoyed the activity and perceive that the practice, particularly shadowing, helped them to improve intonation, stress, pitch, volume, speed, and rhythm, the LLs were also discouraged or dissatisfied with their outcomes, which may be a product of comparing themselves to native-speakers as the model—a major factor to consider in terms of language-learning identity.

Finally, teachers may need to continually revisit the growth in comprehensibility of students for continued motivation and development of proficient-speaker identity. While LLs are becoming consciously aware of gaps between their performance of the TT and the professional speaker's, teachers need to acknowledge that awareness to motivate as well as provide opportunities to meet with LLs' learners goals.

In many ways, the activity suggests identity-shifts in terms of language use for the LLs in helping to notice gaps between their production and the models they chose, but for these advanced students, noticing the gap (Schmidt, 1990) is not enough, as they are also aware that production or shifting the production to be more comprehensible is their ultimate goal. This activity makes LLs more aware of the difficulty of advancing from noticing (awareness) to producing (application). While the activity raises awareness, it may be identity-impeding for LLs in perceiving themselves as proficient speakers of English. What may be more important in getting the students to "notice the gaps" or raise awareness towards comprehensibility is to also raise their awareness in their shift in becoming more comprehensible, which may not necessarily entail native-like production, but more comprehensible production. This may be achieved in two ways. First, choosing comprehensible non-native models or urging the LLs to do so may mitigate the disappointment the LLs seem to be feeling in comparing and contrasting their own production to a native-speaker model. Secondly, using Native-speaking models, but concentrating on growth in comprehensibility for the LL, not just the final product of the presentation may be more beneficial at facilitating LLs in building identities in which they see themselves as gaining comprehensibility, not falling short of the model. Using both of these solutions may also help in facilitating the LLs' understanding comprehensibility but also of language diversity, particularly a growing and ever-expanding global English or as J. M. Murphy argues in his 2014 research, *Intelligible, comprehensible, non-native models in ESL/EFL*

pronunciation teaching, "question[ing] the hegemony of native English speaker (NES) models…."(p. 258).

Future Directions

This exploratory pilot study has raised further questions. It was proposed to merely answer a question about enjoyment and perceived improvement for LLs to decide whether to continue an activity introduced into the program curriculum. However, the following questions arose: Is this a comprehensibility exercise or an identity exercise, or both? I originally chose it as a pronunciation exercise to work on overall comprehensibility. However, after examining the surveys and the interviews, it raises the question of when language teachers ask students to imitate native speakers, are they contradicting the current wave of research that finds *intelligibility* and *comprehensibility* can change without a shift in *accent* (Munro & Derwing, 1995)? Are instructors establishing an unobtainable model if students utilize the native speaker for imitation activities, and how is this type of activity identity shifting, impeding, or both in helping students to perceive themselves as proficient speakers?

In her Association for Applied Linguistics presentation, Ortega (2010) discusses the bilingual turn in SLA, noting that too often "SLA discourses construct L2 learning and learners as defined by impossibility and failure, bounded by deficiency and disadvantage" (slide 63). The discourse of the LLs in this research reflects an internalization of the monolingual bias Ortega discusses, as the LLs' discourse in this research suggests feelings of inauthenticity or inadequacy when comparing themselves to NSs and it also questions the notion of Native-English Speaking (NES) models (Murphy, 2014). While this research sought to test perceptions of an activity, it became clear that the LLs, while enjoying the activity, are navigating language-learning identities, and that this activity induced the LLs to reflect upon that navigation.

More work needs to be conducted here, particularly on whether the activity facilitated the LLs in improving their comprehensibility, which is another study for the future to see if the LLs' perceptions of improvement match both expert and naïve raters' perceptions of improvement. Work on unpacking LLs' understanding of the term *pronunciation* is also paramount in order to help LLs understand what facilitates or impedes their comprehensibility, and particularly understanding learners' goals and understandings of pronunciation will be important for this type of activity's use. Finally, perhaps in conjunction with unpacking the term *pronunciation* for LLs, a further need for the future is in how best to assist LLs' in their development of identities as autonomous, comprehensible, and proficient speakers, so they do not just notice the gaps, but fill them.

ACKNOWLEDGMENTS

ABOUT THE AUTHOR

Janay Crabtree is faculty at the University of Virginia where she directs a summer English for Academic Purposes Program and teaches both linguistics and ESL as well as administrates the TESOL Certificate Program. When she is not contemplating perception of language learners, second language acquisition, language teaching, and language teachers and their training, she enjoys running after her two spritely children, fitness, and cycling. Her M.A. is from Western Kentucky University and her Ph.D. from the University of Georgia. Currently, she is interested in helping learners become autonomous self-evaluators. She can be reached at jc9ne@eservices.virginia.edu.

REFERENCES

Crotty, M. (2003). *The foundation of social research: Meaning and perspective in the*

research process. London: Sage.

de Guerrero, M. C., & Commander, M. (2013). Shadow-reading: Affordances for imitation in the

language classroom. *Language Teaching Research, 17*(4), 433-453.
doi:10.1177/1362168813494125

Goodwin, J. (2007). Mimicking American TA discourse. In Catherine Ross & Jane Dunphy

(Eds.*), Strategies for teaching assistant & international teaching assistant development: Beyond microteaching* (pp. 156-7). San Francisco, Ca.: John Wiley & Sons.

Harmer, J. (2012). *The practice of English language teaching* (4th edition). Edinburgh:

Pearson Longman ELT.

Henrichsen, L. (2014, September). *Video voiceovers for helpful, enjoyable pronunciation*

practice. Teaching tip presented at Pronunciation in Second Language Learning and Teaching Conference, Santa Barbara, CA.

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in

the speech of second language learners. *Language Learning, 45*, 73-97.

Murphy, J. M. (2014). Intelligible, comprehensible, non-native models in

ESL/EFL pronunciation teaching. *System, 42*, 258-269.
doi:http://dx.doi.org.proxy.its.virginia.edu/10.1016/j.system.2013.12.007

Ortega, L. (2010). The bilingual turn in SLA [Powerpoint slides]. Retrieved from

https://sites.google.com/a/georgetown.edu/lourdes-ortega/

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*

11, 129-158

Appendix A

Survey questions for research project: *In Other People's Words: Nonnative-Speakers' Imitation of Professional Speech. If you need extra space, please feel free to write on the back.*

1. Which TED talk did you choose?
2. Why did you choose this particular TED talk?
3. Did you listen to the practice script? If you answer no, skip to question #9.
4. If the answer to #3 is "yes," how many times did you listen to the practice script?
5. If you answered "yes" to # 3, did you practice the TED talk with the video? How many times or for how long?
6. If you practiced the activity, how did you practice? What did you pay attention to as you practiced?
7. If you practiced the activity, did you mute the speaker or did you follow right after the speaker?
   Describe your method of practice.
8. If you practiced the activity, how did you feel as you voiced over your speaker?
9. If you did not practice the activity, how did you feel as you performed the talk for the class?
10. Do you feel you improved your pronunciation from the practice script to the final presentation? If so, how?
11. If you answered "yes" to #8, please circle on the scale below how much you feel you improved with 1 being "no improvement" to 5 being the "greatest improvement:"
    1   2   3   4   5
    No Improvement   Greatest Improvement
12. What are your thoughts about the activity? Would you like to take part in a similar activity again? Why or why not?
13. Rate your experience with this activity in terms of enjoyment with 1 being "did not enjoy at all" and 5 being "enjoyed a great deal."
    1   2   3   4   5
    Did not enjoy at all   Enjoyed a great deal
14. Did you review your voice-over final TED presentation provided for you in Kultura in our collab site? If so, how many times?
15. If you did review your presentation, how did you feel as you watched your voice coming from the TED speaker? What adjectives would you use to describe how you felt as you watched and listened to your voice over?
16. If you did not review your presentation, why not?
17. Does this activity help you view yourself as a proficient English speaker? Why or why not?

Other Comments:

Appendix B

Table 2: Similarly themed statements of dissatisfaction made by LLs

Language Learner	Statement of Comparison of Voiceover to the Professional speaker
KYJ	I think it's a good step, and I really like this imitation process, but I will have some concerns about how [we can] use it in our daily talking… I still speak the old way in daily talking.
KYL	Though I felt helped by practicing the speech, I realized it is still a long way to go to be a proficient English speaker to address a speech precisely.
KJ	Yes, it's a good way for me to see the gap between different levels of English speakers, and it enable[d] me to see the progress I've made by practicing often.
JZ	{On reviewing presentation & viewing one's self as a proficient speaker} I think for the TED talk, my performance is not that good, not just because of pronunciation. I am also not satisfied with the speed and just trying to catch the speaker…& [I] did it in a monotone way…
JHH	{On proficiency} Especially because I realize I don't stress words a lot like you guys do, yeah, it was helpful with stress, but I kind of notice how Americans are speaking, American pattern.
KLW	{On pronunciation improvement} I think so. At first, I can't catch up [with] the speed of the lecturer. {On reviewing presentation} I feel a little bit awkward. {On proficiency} Yes, this activity makes me to view myself as a proficient English speaker. If I work hard on one thing, I can perform well.
JG	{On practicing and performing} I recorded the practice and listened to my own voice, and I think I am a little concerned about the fluency of my talk…when I listen to my voice, I can feel the difference…maybe stress or something like that.
JB	{On reviewing presentation} My voice is dim, and I used to think that I would do the presentation, I would try to sound really clear, after, I see the videos, I feel a little disappointed [in] myself.
KFW	{On reviewing presentation} It sounds a little bit unnatural, funny yet awkward. {On proficiency} It has helped me obtain more insights of how native speakers speak, which inspires me to practice more and follow.
KNW	I sound less emotional than I intended to. Unsatisfied…when I really hear myself talking in the video, I still don't think I talk like a native, even if I tried so hard to mock the TED professor.
JY	{On reviewing presentation} No, I did not watch… I don't want to see what I am saying. I can not say why I don't want to. It just makes me uncomfortable.

A PRACTITIONER'S GUIDE TO ENGLISH RHYTHM: A RETURN TO CONFIDENCE

Wayne Dickerson, University of Illinois at Urbana-Champaign

The TESOL version of English rhythm, often called stress-timed rhythm, comes to us in a coherent and persuasive narrative, honed by decades of unquestioned acceptance and use. Appearing in the 1950s, this version enjoyed 30 years of near-universal popularity within the profession. However, during the latter half of this period, linguistic researchers found a uniform lack of empirical support for the core tenets of this model. By the early 1980s, evidence against the model became too great to ignore, launching a period of growing doubt among TESOL professionals about how to describe English rhythm. After more than three decades in this unsettled state, we can now see beyond stress timing to an alternative model of rhythm and to a return to confidence about how English rhythm works. This guide traces the history of our growth and assembles the critical evidence underlying it. The intent is to make it easier for ESL/EFL teachers and teacher educators to describe and to teach English rhythm. It also cautions practitioners about continuing to promote TESOL's now-discredited model.

INTRODUCTION

English rhythm has always been described in hedging language. For example, note the words *tend to, relatively, approximately,* and *in general* in the following.

> In sentence rhythm the stressed syllables tend to occur at relatively regular intervals.... This uniformity is preserved when the number of syllables in each rhythm group varies; but each group occupies approximately the same amount of time (Fries, 1943, p. 200).[1] In general *content words* are *stressed*, but *function words* are left *unstressed* (Prator, 1951, pp. 25-26 [italics in the original]).

This mild equivocation was acceptable because stress-timed rhythm was assumed to be a dominant, if not categorical, speech behavior in English. Now, more than 60 years later, and after this assumption has been found to be false, the tone is different. Having described stress timing essentially as above, Celce-Murcia, Brinton, Goodwin, and Griner (2010) continue:

> We should note here that the distinction between stress-timed and syllable-timed languages is not universally accepted.... However, most pronunciation researchers

---

[1] In Charles Fries' 1943 ESL materials, Kenneth Pike prepared all the pronunciation lessons (Fries, 1943, p. iv). When citing Fries (1943), we are referring to Pike's contribution in all cases.

and practitioners agree that stress-timing represents at least a strong tendency in English and is thus critical to include in the pronunciation curriculum. (p. 208)

What happened between the sweeping *in general* of the 1950s and the more hesitant *at least a strong tendency* in 2010 to shake our confidence about a fundamental part of English phonology? How should we describe English rhythm? What should practitioners be teaching now? This guide has been written to answer these questions and clarify the direction forward.

A TIME OF CONFIDENCE

Two key figures, both ESL teachers, had a profound impact on TESOL's model of rhythm in the middle of the last century. Kenneth Pike was a professor and gifted linguist at the University of Michigan, doing seminal work on English intonation and rhythm. Clifford Prator, Jr. was a well-respected professor and ESL professional at UCLA, and an effective popularizer.

The model of rhythm that arose from their work was so simple and teachable, and was stated with such authority, that it settled the matter of how to describe English rhythm for most ESL textbook writers and teachers. Its three supporting pillars can be labeled succinctly.

The first, **stress alternation**, states that "the alternate stressed and unstressed syllables and the alternate high and low pitches form a *sentence rhythm*" (Fries, 1943, p. 200 [original italics]).

The second pillar concerns the timing of heavy stresses, or accents. As noted above, accents tend to recur at regular intervals so that the time between the heavy stresses—called the **interstress interval**—is about the same from phrase to phrase. This is Pike's contribution to the model. Although not the first to assert the regularity of heavy stresses (see Jones, 1918, p. 106), Pike is remembered best for this pillar because of naming it **stress-timed rhythm** (Pike, 1945b, p. 35).

The third pillar identifies where these heavy stresses occur in phrases: Every content word carries an accent.[2] This rule first appears in TESOL literature in Prator's 1951 pronunciation textbook entitled *Manual of American English Pronunciation for Adult Foreign Students*.

To trace the source of Prator's claim, we checked to see if it might be Pike. Prator and Pike were not only contemporaries, but Prator also drew on Pike's work extensively as he developed his pronunciation textbook.[3] On investigating, we learn that the accent-every-content-word rule is contrary to Pike's research. Our conclusion is that Prator himself is the source. It seems likely

---

[2] Pike defined **content words** as nouns, adjectives, adverbs of time, place, and manner, verbs, interrogative words, demonstrative and indefinite pronouns, and interjections. The rest are **function words** (Pike, 1945a, p. 118). An **accent** is a heavy stress with a pitch change.

[3] Prator does not cite Pike as the source of his description of rhythm's timing feature. Even so, it is obvious. Prator (a) uses Pike's terms **content word** and **function word,** defined as Pike does (Prator, 1951, p. 26); (b) borrows Pike's reference to Tennyson's poem; compare Pike (1945, p. 34) and Prator (1951, pp. 24-25); (c) explains stress timing (without using the term) with the same phraseology as Pike; compare Pike (1945, p. 34) with Prator (1951, p. 24); (d) rewords two of Pike's examples to make the same point; compare Pike (1945, p. 34) with Prator (1951, p. 25).

that when Prator read Pike's rules, Prator felt they were too complex for his ESL students to use and taught instead that *every* content word should carry a heavy stress.

Prator's textbook was so easily accessible that it quickly became the most popular American pronunciation text in our profession, going through four editions (1951-1985). Along the way, it taught generations of TESOL practitioners about English rhythm, this author included. With a measure of overconfidence, authors repeated the three pillars of the model in virtually every pronunciation text since.[4] Lacking serious competitors, Prator's description of rhythm became TESOL's model and is now known worldwide.

A TIME OF UNCERTAINTY

The seductive elegance of Prator's model and the profession's enthusiasm for it could not hide the fact that its foundations were untested hypotheses. When linguistic researchers examined the model, they exposed how serious the cracks were in some of the pillars assumed to support this conceptual edifice.
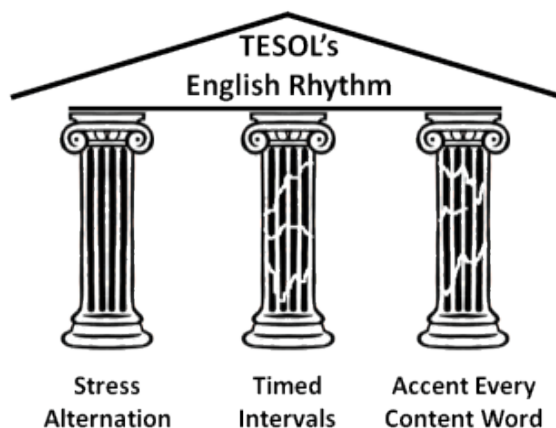
Stress alternation. The first pillar—the presence of large swings in the prominence of syllables across a phrase—is incontrovertible.

Pike's timing pillar. The second pillar, dealing with timing of accents, has catastrophic cracks. Arvaniti (2012, pp. 351-353) summarizes the many studies that definitively demonstrate that interstress intervals in English are not uniform in length but are proportional to the number of interstress syllables. The timing of heavy stresses is so variable that it contradicts the claim "that stress-timing represents at least a strong tendency in English" (Celce-Murcia, et al., 2010, p. 208). The regular timing of accents is therefore not a pillar of English rhythm.

---

[4] Some of those who include Prator's description of English rhythm are Beisbier, 1994, pp. 35-52; Chan, 1987, p. 113; Dale & Poms, 1994, p. 224; Dauer, 1993, pp. 84-87; Gilbert, 1984, pp. 34, 37; Grant, 2001, pp. 85-86; Hahn & Dickerson, 1999, pp. 43-46; Kozyrev & Stein, 2001, p. 27; Lane, 2005, p. 156; Miller, 2006, pp. 76-82; Zawadzki, 1994, pp. 141-144.

Prator's accent pillar. The third pillar embodies the claim that we "accent every content word." It is not only foreign to Pike's research but also to his pronunciation teaching.[5] Throughout his teaching materials, Pike insists that learners speak in a spontaneous conversational style of speech which he describes this way: "A conversational style is characterized by *few centers of special attention* [accents] and by *many repressed lexical stresses"* (Pike, 1945a, p. 72 [Pike's emphasis]). To emphasize the second point, he says, "With extreme frequency word stress is completely suppressed in context" (Pike, 1942, p. 31). He then lists nine contexts in which the lexical stress of content words is demoted, for example, "Lower the stress marks between any two syllables with special attention [accents], within a rhythm unit" (Pike, 1945a, p. 65). Bullets in these sentences mark Pike's accents:

○   ●

The new doctor's not a very good student. (Fries, 1943, p. 203)

Here the stress of *not, very,* and *good* is downgraded. Next, the stress of *going away* is downgraded.

○   ●

She told me that Emily was going away to boarding school. (Pike, 1945a. p. 65)

Pike's other suppression rules account for the lowering of stress of content words before and after the principal accents, such as *new, told,* and *school* above.

Putting his observation about stress suppression into practice, Pike instructs teachers and students on how to use his pronunciation exercises as follows:

> Pronounce the following sentences rapidly and evenly.... Observe the suppression of normal stress on some of the nouns, verbs, adjectives, and adverbs in this rapid pronunciation. (Fries, 1943, p. 292)

From these examples, Pike's position is clear: English speakers do not accent every content word. In fact, Pike's research effectively destroys the accent-every-content-word pillar.

To his credit, Prator knew his rule was not entirely adequate and struggled to improve it in the early editions of his textbook. Echoing Pike's observation that native speakers regularly downgrade the normal stress of content words, Prator warned his students that native speakers will violate the accent-every-content-word rule: "A native speaker of English might feel this to be an unnatural rhythm and instinctively suppress some of the stresses" (Prator, 1957, p. 27). If the rule yields an unnatural rhythm, why does Prator persist in using it in all subsequent editions of his textbook? One reason may have been that the rule makes it "a simple matter to determine

---

[5] It is important to be clear: What linguists called stress-timed rhythm and what TESOL professionals called stress-timed rhythm were not the same thing. TESOL's version included Prator's accent-every-content-word rule. This rule was absent from the linguistic version.

where the stresses are placed in a sentence" (Prator, 1972, p. 33). Also, "The basic principles—content words stressed, function words unstressed—are easy to follow" (Prator, 1972, p. 34).

For many TESOL professionals responsible for preparing the next generation of teachers and researchers in the field, the revelation that English rhythm is not stress timed was unsettling. When they asked: *What kind of rhythm does English have then?* their fundamental question was, *If not stress timed, what kind of timing, if any, does English have?* Celce-Murcia et al. (2010) above address the timing issue: (a) Keep teaching stress-timed rhythm because it is surely a strong tendency in English, and (b) note that some scholars doubt the validity of stress timing.

At the same time, Prator's accent rule was also under attack. Linguists did not challenge the accent-every-content-word rule directly because it had no place in their rhythm model. Their indirect challenge, however, surfaces in the examples they use and the comments they make in their research, as shown in the citations from Pike above (and below).[6] TESOL professionals were largely unaware that linguists accepted stress suppression as normal in English.

The fact is that linguistic research swept away *both* key tenets of TESOL's rhythm model, not just one. By focusing on the timing pillar and ignoring the accent pillar, TESOL professionals gave their tacit approval to continue teaching that every content word should be accented, even though the distortion it creates is more serious than the distortion of the timing pillar.[7]

A RETURN TO CONFIDENCE

In order for TESOL practitioners to return to a comfortable level of confidence about English rhythm, two needs must be met: (a) an acceptance that accents are not regular in English, and (b) a proven alternative to the stress-every-content-word rule. For the latter, we return to Pike. Although his emphasis on regular timing was not justified, he offers something different for the accent pillar and deserves credit for an insight that points the way beyond accenting every content word.

Throughout his writing on English rhythm, Pike provides many examples of a rhythm that he describes this way: "Usually only one or two syllables within a rhythm unit (that is, between two pauses) will receive special attention [an accent]" (Pike, 1945a, p. 64). The following are phrases having one or two accents and from 0 to 4 content words with suppressed stresses.[8]

---

[6] Wells (2006) is one of many researchers whose comments and examples also illustrate the acceptance by phoneticians of stress suppression: "The option to downgrade potential accents is a pervasive characteristic of English rhythm" (229).

[7] It can be demonstrated that speakers can compress multiple interstress syllables to keep accents regular (timing pillar), even when they do not do so in everyday speech. As the last section below shows, multiple accents (accent pillar) can cause serious communication damage.

[8] The following eight examples from Pike also illustrate the neutral anchor-placement protocol in full (Dickerson, 2015).

○   ●   ○   ●

One man is here.   Mrs. White's little boy went to the house.

   Pike, 1945b, p. 75   Fries, 1943, p. 359

       ○   ●   ○   ●

He seldom comes early.   Buy me some big brown potatoes.

   Fries, 1943, p. 151   Pike, 1942, p. 62

       ○   ●   ○   ●

None of them believed it.   How does he get along with the students?

   Fries, 1943, p. 340   Fries, 1943, p. 731

              ○   ●   ●

Could you tell me the time?   Do you remember him?

       Fries, 1943, p. 664   Fries, 1943, p. 221

Pike is not alone in noticing that the dominant rhythm patterns of spoken English have only one or two accents.[9] The names phoneticians commonly associate with these accents are the **onset** (the hollow bullets above) and the required **nucleus** (the filled bullet above). Given the results of later research, linguists now make no claim that these accents occur at regular intervals.

Pike's accent rule for English is so well documented in linguistics that we (Laura Hahn and this author) built the new edition of our pronunciation text, *Speechcraft,* around it (Dickerson & Hahn, forthcoming). We use the metaphor of a mountain range in silhouette and call this model of English rhythm the **two-peak profile**. We refer to the first peak as the **anchor peak** or just the **anchor** and to the second (or only) peak as the **primary peak**. The valleys—before, between, and after these peaks—consist of unstressed or weakly stressed syllables, including the suppressed stresses of content words. See Dickerson (2015) for the neutral anchor-placement rule.

---

[9] In Pike's study of read-aloud conversations, 86.1% of all accented phrases had only one or two accents (Pike, 1945b, p. 151). This finding is similar to that of Cauldwell who found that 91% of phrases in spontaneous speech had only one or two accents (Cauldwell, 2002, p. 15). Bolinger (1961), Brazil (1980), and Wells (2006) make similar claims about unrehearsed speech. While it is encouraging to find the same patterns in British and American English, we look forward to the results of phonetic analyses of large corpora that show us the relative prevalence of these rhythm patterns in other varieties of English of interest to sociolinguists.
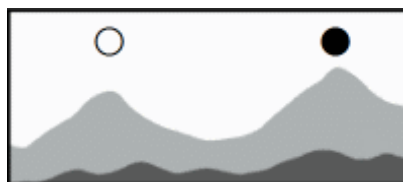
Although the two-peak profile has been known for at least 70 years, starting with Pike, and has been repeatedly confirmed as characteristic of spontaneous speech, its viability for pedagogical purposes has been limited. Until recently, the position of the anchor had not been well defined nor cast in learner-friendly language. Our work to address these problems has made the two-peak profile available as a replacement for Prator's accent rule in TESOL's rhythm model and as a means to restore a sense of confidence to TESOL professionals about their understanding of English rhythm.

THE COST OF INDECISION

With an alternative to accent-every-content-word in hand, there are now some compelling reasons to bring the two-peak profile explicitly into certain TESOL classrooms and some non-trivial costs to students if instructors do not. To illustrate, we cite Pike himself. These reasons and costs are relevant primarily to learners who want to interact easily with speakers of native varieties of English such as North American English, British English, Australian English, New Zealand English. This is because the expectations of such listeners shape speakers' requirements.

The **first cost** to learners of accenting every content word is that their speech does not sound natural to native users of English, as Prator admits (1957, p. 27). No native speakers of English (of the varieties noted above) speak acceptably if they stress every content word in every phrase.

    Unnatural: The manager doesn't often pay his bills. (Fries, 1943, p. 292)

Saying phrases this way, speakers will certainly be accused of speaking with an odd accent or affecting an imitation of a non-native variety of English.

Pike does not equivocate: He calls stress on every content word "slow speech" and "bad accent" (Fries, 1943, p. 102). When his students spoke this way, he gave them rhythm exercises emphasizing suppressed stresses and a faster pace. One of his exercise items is this same sentence which he marks with accents indicated by the bullets.

    Natural: The manager doesn't often pay his bills. (Fries, 1943, p. 292)

Commenting on the effect of these stress-suppression and speed exercises, Pike says, "It was then observed that when the rhythm, speed, and grouping of syllables were correct, that objectionable unnaturalness disappeared" (1945b, p. 109).

Since the two-peak profile matches what native speakers do naturally when they speak, native English teachers can readily model it for their students. The better they follow this model, the more comprehensible their speech is to native-English listeners.

A **second cost** to learners of stressing every content word is that their speech is not as polite as they may want it to be when speaking with native English listeners. A native speaker of English may occasionally accent every content word in a phrase for emphasis. It is unnatural for a native speaker to say *every* phrase this way because of the way it will be interpreted. What is the emotional impact on a native English listener of a conversation filled with what Pike calls "emphatic" phrases like these (Fries, 1943, p. 174)?

○   ○   ○   ●

He never had a chance to apologize.

○   ○ ○   ●

Or maybe he was just too proud.

The impression is that of insistence. Conscientious students who are trying to implement Prator's rule may sound pushy and rude, even irritated or aggressive. They cannot sound calmly polite to a native listener. By contrast, the two-peak profile registers as neutral with respect to such overtones:

○   ●

He never had a chance to apologize.

○   ●

Or maybe he was just too proud.

A **third cost** to learners of stressing every content word is that their speech can delay native listeners' sentence processing and slow their comprehension. Why do extra peaks have this effect? Gorsuch, Meyers, Pickering, and Griffee (2010, p. 7), following Brazil (1997), correctly observe that native speakers typically highlight one or two salient words in each thought group. If a speaker highlights additional peaks, they compete for attention, creating a problem for native listeners who will naturally try to pick out the main concepts for themselves—something they may find difficult to do (Gorsuch, et al., 2010, p. 26).

Something else may also be at work: slowed speech. Peak vowels have greater duration than vowels in valleys. More peaks take more time, slowing delivery and hurting understanding. Pike noticed this: "Slow speech hinders the comprehension of normal English" (Fries, 1943, p. 102). Munro and Derwing (1998) and others have found the same effect of slowed speech.

The underlying mechanism may have to do with the function of the two accents in the two-peak profile. These two peaks are not on just any two words, but on the two words that *together* give native English listeners a semantic snapshot of a phrase, which other words in the phrase fill out. What is the essence of *He nêver had a chance to apólogize?* It is in the combination *never... apologize*—anchor and primary peaks. What about the essence of *Mâybe he was just too próud?* It is in the word pair *maybe... proud.*[10]

He never had a chance to apologize.
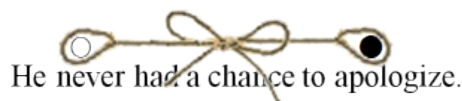
Olle Kjellin (1999, pp. 23-24) says that listeners grasp this essence only if two conditions hold. The first condition: The peaks must be close enough together to register as a single thought. Extra peaks between the anchor peak and primary peak, each with a vowel of longer than average duration, push the anchor and primary peaks apart. If the speaker does not suppress the extra peaks and compress all other valley syllables to shorten the time between the anchor and primary peaks, the native English listener may not hear the main peaks as a unit nor understand the speaker's message so readily.[11]

The second condition: It is not just a matter of intervening time, Kjellin says, but efficient processing "seems to require that the *speech rhythm* be the expected one" [original emphasis]. He continues, "If it's not, perception will work slowly and inefficiently, sometimes not at all" (Kjellin, 1999, p. 24). So the unexpected rhythm caused by extra peaks can also undermine the native English listener's immediate grasp of the message.

By contrast, the two-peak profile exactly accommodates native listeners' processing needs: It allows no intervening peaks and compresses valley syllables to draw the anchor and primary peaks together—condition 1. It does both things with a rhythm that listeners are used to and expect—condition 2. The speaker thereby helps native listeners minimize processing delays and semantic loss.

The **fourth cost** is that TESOL's version of English rhythm cannot easily camouflage other errors in speech; instead, it highlights them (Kjellin, 1999, p. 24). Pike notes this too: "If pitch and stress are correct, slight errors in sound will not be so prominent" (Fries, 1943, vol. 1, p. vii). This is another bonus of the two-peak profile. But when the rhythm itself is unexpected—as

---

[10] In descriptions of English phonology, **prominence,** i.e., tonic or focal stress, is understood to be an important part of prosody (e.g., Celce-Murcia, et al., 2010, pp. 223-225). We expand the term to include the other accent, the anchor peak, because it and the primary peak work together as a unit. Brazil, et al. (1980, 42, 45), from the perspective of discourse intonation, also consider the two as a single unit, the **tonic element**, which carries all the intonational meaning of a tone unit (message unit).

[11] The **contrast** dimension of rhythm has a *meaning*-based motivation, namely, to highlight the most important words (Pike, 1945a, p. 73). The **compression** dimension had a *timing*-based motivation, that is, to keep interstress intervals the same size (Fries, 1943, p. 291; Pike, 1945b, p. 34). Since regular timing does not exist, why squeeze valley syllables? We do so for a *meaning*-based reason, namely, to draw the anchor and primary peaks together so the listener hears them as a unit, as the gist of the message. Meaning is the rationale for both dimensions of rhythm.

when accenting every content word—native English listeners attend to every clue in the phrase to make sense of what is said. In this heightened state of noticing, other errors—segmental errors—stand out along with the prosodic ones.

CONCLUSION

The journey our field has taken to grow in its understanding of English rhythm has at times been difficult. It began with the impact of two practitioners of the last century, each with his own pronunciation teaching materials and model of rhythm. While both were alike in the area of timing, each was unique in the area of accenting.

At a critical fork in the road, Clifford Prator's accent lessons oversimplified English rhythm. His rule to accent every content word in a phrase went viral and became part of TESOL's model of rhythm. At the same fork, Kenneth Pike's accent lessons introduced a model of rhythm with at most only one or two accents per phrase. By contrast, his lessons aroused little interest. Today we are back at the same fork in the road because years ago our field preferred the simple, easy-to-use rhythm rule that unfortunately misrepresents English and can compromise the acceptability and intelligibility of students' speech to native English listeners. With the clarity of hindsight and the guidance of researchers like Pike, Bolinger, Brazil, Cauldwell, and Wells, we should be able to choose the road not taken before and be confident that it represents an authentic version of English rhythm that will benefit speakers and native listeners alike.

ACKNOWLEDGEMENTS

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), University of Illinois Press, and (with co-author Laura D. Hahn) *Speechcraft: Discourse Pronunciation for Advanced Learners* (1999), and associated workbooks, The University of Michigan Press.

REFERENCES

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

Beisbier, B. (1995). *Sounds great, intermediate pronunciation and speaking for learners of English, book 2.* Boston: Heinle & Heinle.

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

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

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

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

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.

Chan, M. (1987). *Phrase by phrase: Pronunciation and listening in American English.* Englewood Cliffs, NJ: Prentice-Hall Regents.

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

Dauer, R. (1993). *Accurate English, a complete course in pronunciation.* Englewood Cliffs, NJ: Regents/Prentice Hall.

Dickerson, W. (1989/2004). *Stress in the speech stream: The rhythm of spoken English.* Urbana, IL: University of Illinois Press. Published on CD in 2004 by the author.

Dickerson, W. (2015). A NAIL in the coffin of stress-timed rhythm. In J. Levis, R. Mohamed, M. Qian & Z. Zhou (Eds.). *Proceedings of the 6th Pronunciation in Second Language Learning and Teaching Conference* (ISSN 2380-9566), Santa Barbara, CA (pp. 183-193). Ames, IA: Iowa State University. Available at: https://apling.engl.iastate.edu/alt-content/uploads/2015/05/PSLLT_6th_Proceedings_2014.pdf

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.

Gilbert, J. (1984). *Clear speech, pronunciation and listening comprehension in American English.* Cambridge, UK: Cambridge University Press.

Gorsuch, G., Meyers, C., Pickering, L., & Griffee, D. (2010). *English communication for international teaching assistants.* Long Grove, IL: Waveland Press.

Grant, L. (1993). *Well said, advanced English pronunciation.* Boston: Heinle & Heinle.

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

Jones, D. (1918). *An outline of English phonetics.* Leipzig and Berlin: B. G. Teubner.

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

Kozyrev, J. R., & Stein, M. B. (2001). *Talk it through! Listening, speaking, and pronunciation 2.* Boston: Houghton Mifflin.

Lane, L. (2005). *Focus on pronunciation 3.* White Plains, NY: Longman.

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

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

Pike, K. (1942). Pronunciation. In *Intensive course in English for Latin-American students.* Vol 1. Ann Arbor: The University of Michigan Press.

Pike, K. (1945a). Appendix: Step-by-step procedure for marking limited intonation with its related features of pause, stress, and rhythm. In *Teaching & learning English as a foreign language,* C. Fries. Ann Arbor: The University of Michigan Press.

Pike, K. (1945b). *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.* Los Angeles: University of California Press.

Prator, C., Jr. (1957). *Manual of American English pronunciation.* New York: Rinehart & Company.

Prator, C., Jr., & Robinett, B. W. (1972). *Manual of American English pronunciation* (3rd ed.). New York: Holt, Rinehart and Winston.

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.

WHERE DOES PRONUNCIATION STAND IN THE 21ST CENTURY FOREIGN LANGUAGE CLASSROOM? EDUCATORS' AND LEARNERS' VIEWS

Frédérique Grim, Colorado State University

Jessica L. Sturm, Purdue University

Much research has been conducted on the efficacy of pronunciation instruction, most suggesting that pedagogical intervention leads to improved pronunciation. But what are the actual practices and current views of instructors and learners? In order to understand where educators and learners currently stand, this study will first present views on the role of pronunciation in language learning from post-secondary foreign language (L2) instructors and students, gathered through a national survey. Results indicate that students value pronunciation instruction more highly than their instructors. Conclusions drawn from the data will be accompanied by suggestions for further research and pedagogical recommendations.

INTRODUCTION

The first National Standard for Foreign Language Learning calls for *communication*[1], which is defined as the ability to converse, to understand, to interpret and to share coherent information. We teach learners to speak, to listen, to read and to write, ensuring that their goal of successful communication is achieved through a multitude of sociocultural and linguistic cues. Successful communication depends on intelligible pronunciation (Derwing & Munro, 2009). Without intelligible pronunciation, a grammatically-correct sentence becomes less comprehensible and may contribute to a communication breakdown.

Numerous researchers (Dansereau, 1995; Hannahs, 2007; Lord, 2005; Pennington & Richards, 1986; Saalfeld, 2011; Sturm 2013a; 2013b) have noted the lack of pronunciation instruction in the foreign language (L2) classroom. Could pronunciation be following the path of grammar instruction within the communicative approach? First, grammar teaching received a considerable amount of attention in L2 teaching (focus on forms). Next, it was considered unnecessary (focus on meaning). Finally, a more balanced approach used a focus on meaning while still focusing on grammar with communicative goals (focus on form) (Doughty & Williams, 1998). Pronunciation, it seems, is following a similar path, and while it is receiving more attention, might still lack a necessary focus.

Review of Previous Literature

Previous studies concerning a focus on pronunciation attempted to make suggestions within the realm of communicative teaching (Anderson-Hsieh, 1989; Cant, 1976; Denbow, 1994; Elliott, 1997). Additional studies (e.g., Miller, 2012; Lord, 2005, 2008, 2010; Rossiter, Derwing, Maintinim, & Thomson, 2010; Saalfeld, 2011; Saito, 2011; Shively, 2008; Sturm, 2013a; 2013b; Thomson & Isaacs, 2009; Trofimovich & Gatbonton, 2006; Vokic, 2010) have examined the impact of pronunciation instruction and have generally found such pedagogical intervention to be effective. However, as Pennington and Richards (1986) mentioned, pronunciation has traditionally been disconnected from communicative intents and is more often associated with linguistic characteristics; it is therefore important to keep pronunciation imbedded in communicative competence as intelligibility of sounds are an important part of oral communication (p. 208).

Dansereau (1995) asserts that pronunciation should be integrated in beginning and intermediate levels, despite the fact that communicative language approaches have set it aside; pronunciation should be an inherent part for successful communication. In the past 20 years, little has changed. While numerous empirical studies on the effects of pronunciation instruction have been done in the past decade, as listed above, Dansereau (1995) and Pennington and Richards' (1986) statements still seem current, as there is a disconnect between communicative-focused teaching and pronunciation instruction.

Among the various aspects of L2 learning, the skills involved in pronunciation tend to occupy the least amount of time, in or out of class, in language courses (Lord, 2005; Saalfeld, 2011). Instructors do not know, or believe that they don't know, how to teach pronunciation (Derwing & Munro, 2009); current, communicative methodology and practices discourage mechanical drills, such as the production and perception drills associated with pronunciation and phonetics instruction (Saalfeld, 2011); and considerable research on the Critical Period Hypothesis (Bongaerts, et al., 1998; Flege, Yeni-Komshian & Liu, 1999; Walsh & Diller, 1981), along with anecdotal evidence, suggests that L2 learners rarely, if ever, achieve native-like pronunciation.

Rather than native-like perfection, Derwing and Munro (2009) emphasize that intelligibility should be the goal for L2 pronunciation instruction and research. According to Richards (1997), pronunciation is one of the most significant factors to successful communication. If the pronunciation of an individual is impaired, despite perfect grammar, the message might not be understood. In L2 French, for example, L1 American English learners have difficulty pronouncing /y/ and /u/ as two distinct vowels (Darcy et al., 2012; Rochet, 1995; Simon, Chambless, & Alves, 2010). Sturm (2013a) cites the confusion that can arise with the lack of distinction between /y/ and /u/: "au-dessous" [od.su] (underneath) and "au-dessus" [od.sy] (on top of), among numerous minimal pairs differing only in the vowels /y/ and /u/. Another example might happen with the use of the non-inverted question, in French. "Ca va?" with rising intonation means "How are

you?", but with a flat or descending intonation, it will likely be perceived as a declarative statement (I am fine!), which would be pragmatically inappropriate in some situations.

Concerned for their intelligibility, students also believe pronunciation has a major role in language instruction (Derwing & Munro, 2009; Harlow & Muyskens, 1994). Derwing and Munro (2009) contend, furthermore, that pronunciation is absolutely learnable and that students do want to learn. Both Derwing and Munro (2009) and Levis (2005) note the supposed "accent-reduction" industry in North America as an indication of ESL students' interest in improving their pronunciation. Harlow and Muyskens (1994) surveyed nearly 1400 L2 French and Spanish learners on their goals for learning their L2. They found that speaking was students' top goal and that pronunciation was ranked fifth out of fourteen stated goals.

While Berri (2000), Calaque (1981), Dansereau (1995), and Sturm (2013a; 2013b) agree that pronunciation needs early focus, beginning-level textbooks still offer only small pronunciation snapshots, covered quickly at best (Grim, 2016; Pedoya, 1984). Many universities offer "Phonetics and Pronunciation" courses typically taken in the third year of coursework, as discussed in Lord (2005), Saalfeld (2011), and Sturm (2013a; 2013b). Miller (2012), however, incorporated pronunciation instruction in a beginning French course.

In the studies by Lord, Saalfeld, and Sturm, participants in the experimental groups were all enrolled in an advanced-level course dedicated solely to Spanish or French phonetics and pronunciation. These students were compared to students enrolled in other advanced-language courses. Lord (2005) and Sturm (2013a; 2013b) found an effect for instruction while Saalfeld (2011) found a ceiling effect. On the other hand, Miller's (2012) participants were enrolled in French 101 (first-semester); she compared two instructional techniques and found that IPA-based instruction was more effective (and more popular with students) than rhyming or sound-alike words as pronunciation guide. Miller received feedback from the French learners, requesting more explicit instruction and feedback on pronunciation accuracy. They preferred being aware of their mispronunciation and using IPA as a means to clearly state what they needed to focus on.

While much research has been and continues to be devoted to the efficacy of pronunciation instruction, not much investigation into learners' (Harlow and Muyskens, 1994 being an exception) and, even more so, teachers' perceptions of the place of pronunciation instruction in the L2 classroom. Researchers have been preaching to the converted on the virtues and necessity of teaching pronunciation. Yet we know little about how teachers feel about teaching pronunciation, and how teachers' attitudes compare to students.

Research question

Due to a lack of investigation on teachers' and students' perceptions and comparisons of those perceptions, this study poses the following question:

1. Do student and teacher perceptions of the role of pronunciation in the communicative language classroom align?

METHODS

Questionnaire

A series of questions (Appendices A and B) was created to elicit educators' and students' opinion on pronunciation. The questionnaires were distributed through several platforms (language departments, ACTFL blog, regional language organizations, and Facebook).

The questionnaire was designed to indirectly assess teacher and student attitudes toward L2 pronunciation by asking open-ended questions about teaching/learning priorities. In this way, we sought to solicit honest, gut-reaction responses to our questions. By not asking specifically about pronunciation, we were able to gauge pronunciation's importance to teachers and students: those who did not mention it clearly did not prioritize it.

Participants

Completing the study were 292 students and 57 instructors (total: 349). Languages represented by the students and teachers surveyed were Arabic, Chinese, Classical Greek, English as a second language (ESL), French, German, Italian, Japanese, Korean, Latin, Russian, and Spanish.

RESULTS

We will compare instructor and student answers for Questions 1-4. Figures 1-3 and Table 1 below reflect percentage of teachers and students naming each answer to the question.
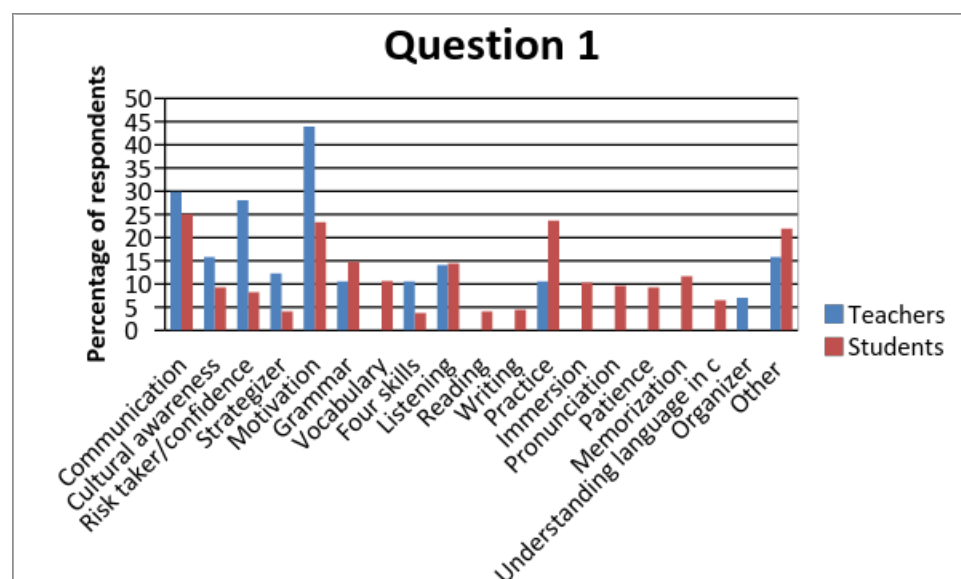
Figure 1. Teachers' and students' answers to the question *According to you, what are the most essential skills in knowing a foreign language?*

Overwhelmingly, teachers named "communication" and "motivation," followed closely by "risk taker/confidence" as the skills that are most essential in knowing a foreign language. From these results, we can conclude that instructors do not view pronunciation as an important skill, possibly because they group pronunciation under communication. Some of the comments left by the instructors show that pronunciation was needed for successful communication; however, they do not seem to view it as a fundamental skill to acquire.

Students named motivation, communication, and practice as the most essential skills in knowing a foreign language. In this they are fairly in sync with the teachers. However, 28 students also named pronunciation as one of the most essential skills (compared to only one teacher). This suggests that students are more concerned with pronunciation than their teachers a) realize and b) respond to.
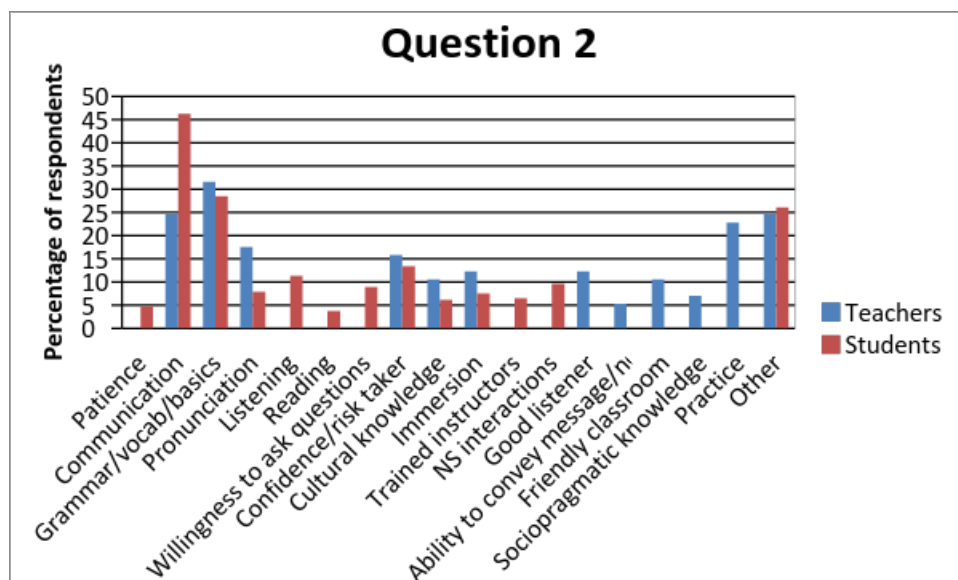
Figure 2. Teachers' and students' answers to the question *"What do students need to achieve adequate/successful communication"?*

Teachers named "communication/authentic discourse/opportunities to speak" along with "adequate knowledge of grammar and vocab," "pronunciation," and "willingness to take risks/confidence" as the greatest necessities to achieving successful communication. Communication is clearly a theme among the teachers' answers, reflecting current communicative methodology and the ACTFL Standards (The National Standards Collaborative Board., 2015) However, in answering Q2, 10 of 57 teachers also recognize the role pronunciation plays in effective language learning.

Students named communication and grammar/vocabulary basics most often as the skills needed for adequate/successful communication, along with confidence and pronunciation. Students believe pronunciation is important to their language learning. Yet they don't mention motivation, which was an important factor according to teachers. It might be more difficult for students to assess their personal motivation and the need for it to succeed in a language. Teachers have a different perspective as they see the impact of a lack of motivation. Students expect more tangible skills, such as pronunciation, to make a difference.

Table 1.

Average rankings from answers to "Rank the following components, in order to importance (1 – most important; 5 least important): Grammar – Communication – Culture – Pronunciation – Vocabulary"

	Grammar	Communication	Culture	Pronunciation	Vocabulary
Teachers	3.18	1.42	3.56	4.25	2.60
Students	3.28	2.18	3.71	3.12	2.67

Both teachers and students ranked communication as the most important component in language learning, followed by vocabulary in second place. After that, the two groups diverge. Teachers rank grammar 3rd; students rank it 4th. Students rank pronunciation 3rd; teachers rank it last. Teachers rank culture 4th but students rank it as least important. Teachers see pronunciation as least important, while students value it more highly in their language learning. This supports their initial answers, that teachers do not view it as an important skill for success while students, in contrast, do.
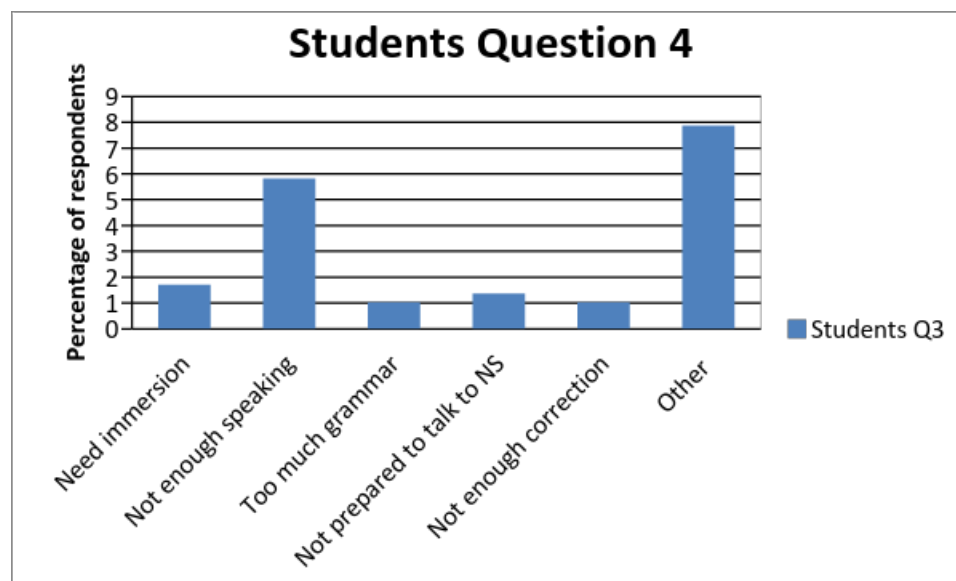
Figure 3. Students' answers to the question *"Do you believe your class(es) prepare them (you) for adequate/successful communication? Justify."*

This question gauged teachers' and students' satisfaction with classroom teaching. The results show that 96% of teachers seem satisfied with either their classes or the curriculum. On the other hand, 82% percent of students felt that they are prepared by their classes for adequate/successful communication.

Pronunciation in Second Language Learning and Teaching 7

Among those who felt unprepared, answers were varied (answers given by fewer than 10 students were combined into the "Other" category). Yet 17 students who said they were not prepared named "not enough speaking" as a reason. Students in this survey want more speaking time; others named the need for immersion or more correction or less grammar.

DISCUSSION

This study shows that teachers and students generally have similar expectations for what is meant by language success and classroom practices. They focused on communication and oral skills as areas necessary for language learning success. Yet disconnects exist between the groups, particularly in relation to pronunciation. Pronunciation appears to have a secondary position in the L2 classroom. This appears to be due to a lack of training (Derwing & Munro, 2009; Pedoya, 1984;) on the part of the instructors. The fact that pronunciation is valued and ranked more highly by students than by teachers shows that students do see the need to acquire more accurate pronunciation in order to be understood by native speakers.

Additionally, as mentioned by Derwing and Munro (2009), teachers may be anxious about their ability to teach pronunciation. Typically, pronunciation and phonetics training is not included in teacher education. Furthermore, non-native speaker teachers may feel inadequate to teach an aspect of the language that they may not feel that they have mastered themselves.

Perhaps most importantly, teachers and students do not seem to understand each other's priorities. From the answers we received, it seems that teachers' and students' objectives are disparate. While teachers have training in pedagogy and teaching approaches, students know what they want to learn, what they find difficult, and what they want to improve. Additional research projects ought to occur in order to better comprehend student and teacher beliefs and allow for more communication between the groups.

Pedagogical Considerations

Acknowledging the student perception that pronunciation should receive more attention in the classroom with the goal of increasing comprehensibility, we need to better prepare teachers to teach it. To this end, the co-author of this paper has developed a graduate-level pedagogy course on Teaching L2 Pronunciation which is offered regularly at her university.

Another important point made by this study is that instructors and students do not seem to be always on the same page in regards to pedagogical goals. Although instructors' class objectives are normally defined at the start of a semester, students' objectives might not be considered and incorporated in the curriculum. It might be helpful to consider students' priorities in language learning from the start, and in turn, either add it to the curriculum or give rationales for not doing so.

Instructors who believe pronunciation should be an integral part of teaching a language and who have received some training with the IPA, segmental, and suprasegmental features might want to consider introducing them from the start, in beginning courses, to a certain degree. Research does support its effectiveness in students' awareness (Miller, 2012; Pennington & Richards, 1986). Having thematic pronunciation lessons could help learners stay within a communicative framework, while making use of some meaningful drill exercises and other types that help them focus on specific sounds. Sturm (2016), for example, gives examples of lessons on French liaison that are short and easily contextualized within a theme or textbook chapter.

Further research

Future studies could examine the role of textbooks and online platforms in pronunciation instruction. Depending on the support provided, teachers may or may not feel ready to cover more pronunciation challenges with their students. In addition, a study on practices could also be very informative, as the results of a series of observations of high school and college L2 teachers could show how educators approach the issues of pronunciation and how they react to students' mispronunciation. In time, this could inform teacher preparation programs on the training to provide to their teacher candidates.

CONCLUSION

From our survey of post-secondary learners and instructors, we draw the following conclusions. First, pronunciation ought to be taught as an integral part of oral language use, as part of the means for creating meaningful communication,, not merely as an aspect of the oral production of isolated sounds, words and sentences. This is based on the fact that both learners and instructors value communication as part of L2 acquisition, as well as students' demonstrated interest in improving pronunciation. Furthermore, pronunciation forms a natural link to other aspects of language use, such as listening, vocabulary, and grammar; ways of highlighting this interdependence in teaching need to be explored. Additionally, we need to teach instructors how to teach pronunciation and support pronunciation instruction. If, as noted by Derwing and Munro (2009) and Saalfeld (2011), instructors are less than confident in their ability to teach L2 pronunciation, then teacher training needs to change to meet the needs elaborated above.

Communication is first among the National Standards (The National Standards Collaborative Board, 2015) because it is the most critical skill to acquire when learning a foreign language. Comprehensibility is developed from a good understanding of the language functions and structures, but pronunciation should be an integral part of the learning experience as well, as it is crucial for effective communication. On a more practical level, students want to work on their pronunciation, and this is a need that instructors should be prepared to meet.

ACKNOWLEDGEMENTS

ABOUT THE AUTHORS

Frédérique Grim (Ph.D. 2005, University of Illinois at Urbana-Champaign) is Associate Professor of French in the Department of Languages, Literatures and Cultures at Colorado State University. Her main research is on teacher education, teaching methodologies (including of pronunciation), service-learning and L1 vs. L2 use. She co-authored *Sons et sens: La prononciation du français en contexte*. (2013).

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. Her research focuses on classroom SLA, specifically the efficacy of instruction on pronunciation in L2 French.

REFERENCES

Anderson-Hsieh, J. (1989). Approaches toward teaching pronunciation: A brief history. *Cross Currents, 16*(2), 73-78.

Berri, A. (2000). Réflexions et propositions à partir de l'article d'Azoulay-Vicente: "Pour un enseignement intégrer de la phonétique". *Travaux de l'Institut de Phonétique de Strasbourg, 30*, 13-31.

Bongaerts, T., van Summeren, C., Planken, B., & Schils, E. (1998). Age and ultimate attainment in the pronunciation of a foreign language. *Studies in Second Language Acquisition, 19*, 447- 465.

Calaque, E. (1981). La phonétique dans les méthodes de niveau 1. In M. Dabène (Ed.), *Recherches en didactiques du Français*, C.D.F. Grenoble, Switzerland: Université des Langues et Lettres de Grenoble.

Cant, J. P. (1976). Phonetic information and pronunciation: Some theoretical considerations. *International Review of Applied Linguistics in Language Teaching, 14*(3), 298-303.

Dansereau, D. (1995). Phonetics in the beginning and intermediate oral proficiency-oriented French classroom. *The French Review, 68*(4), 638-651.

Darcy, I., Dekydtspotter, R., Sprouse, R. A., Glover, J., Kaden, C., McGuire, M., & Scott, J. H. G. (2012). Direct mapping of acoustics to phonology: On the lexical encoding of front rounded vowels in L1 English- L2 French acquisition. *Second Language Research, 28*(1), 5-40.

Denbow, S. (1994). Teaching French to singers: Issues and objectives. *The French Review, 67*(3), 425-431.

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

Doughty, C., & Williams, J. (1998). *Focus on form in classroom second language acquisition.* Cambridge, UK: Cambridge University Press.

Elliott, A. R. (1997). On the teaching and acquisition of pronunciation within a communicative approach. *Hispania, 80*(1), 95-108.

Flege, J., Yeni-Komshian, G. & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language, 41*, 78–104. doi:10.1006/jmla.1999.2638

Grim, F. (2016). How pronunciation is taught: The case of French textbooks. Manuscript in preparation.

Harlow, L. L., & Muyskens, J. A. (1994). Priorities for intermediate-level language instruction. *The Modern Language Journal, 78*(2), 141-154.

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

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

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*(3), 488-503.

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

Pennington, M., & Richards, J. (1986). Pronunciation revisited. *TESOL Quarterly, 20*(2), 207-225.

Pedoya, E. (1984). Pedagogy of pronunciation and the communicative approach: Where are we in it? *Tranel, 6*, 201-214.

Richards, J. (1997). Reflections on language teaching: What to do about pronunciation. *Zielsprache Englisch*, 27(2), 33-34.

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.

Saalfeld, A. K. (2011). Acquisition of L2 phonology in advanced learners: Does instruction make a difference? In. J. Levis & K. LeVelle (Eds.), *Proceedings of*

the 2nd Pronunciation in Second Language Learning and Teaching Conference (pp. 144-152). Ames, IA: Iowa State University.

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.

Shively, R. (2008). L2 acquisition of [beta], [eth] and [ram's horns] in Spanish: Impact of experience, linguistic environment, and learner variables. *Southwest Journal of Linguistics, 27*(2), 79-114.

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.

Sturm, J. L. (2013a). Explicit phonetics instruction in L2 French: A global analysis of improvement. *System, 41*(3), 654-662.

Sturm, J. L. (2013b). Liaison in L2 French: The effects of instruction. In J. Levis & K. LeVelle (Eds.). *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference* (pp.157-166). Ames, IA: Iowa State University.

Sturm, J. L. (2016). Teaching *liaison* to intermediate learners of French. Accepted for publication in *The French Review.*

The National Standards Collaborative Board. (2015). World-readiness standards for learning languages (4th ed.). Alexandria , VA: Author.

Thomson, R. I., & Isaacs, T. (2009). Within-category variation in L2 English vowel learning. *Canadian Acoustics/Acoustique canadienne, 37*(3), 138-139.

Trofimovich, P, & Gatbonton, E. (2006). Repetition and focus on form in processing L2 Spanish words: Implications for pronunciation instruction. *The Modern Language Journal, 90*, 519-535.

Vokic, G. (2010). L1 allophones in L2 speech production: The case of English learners of Spanish. *Hispania, 93*(3), 430-452.

Walsh, T.M., & Diller, K.C. (1981). Neurolinguistic considerations on the optimum age for second language learning. In K.C. Diller (Ed.), *Individual differences in language learning aptitude* (pp. 3-21). Rowley, MA: Newbury House.

Appendix A

Background information and questionnaire (teachers)

Thank you for being willing to participate in this study on your perceptions as a learner of a foreign language! The following questionnaire should take you between 10 and 15 minutes. If you have any question, please contact the research Dr. Frédérique Grim at Frederique.Grim@colostate.edu.

Name: _____ Age: _____   Gender: M   F

    Circle the appropriate answer(s)

    1. What is the primary foreign language you teach:

Question 1: According to you, what are the most essential skills in knowing a foreign language?

Question 2: What do students need to achieve adequate/successful communication?

Question 3: Rank the following components, in order to importance (1 – most important; 5 least important): Grammar – Communication – Culture – Pronunciation - Vocabulary

Question 4: Do you believe your class(es) prepare them for adequate/successful communication?   Yes   No

    Please, justify your answer.

Appendix B

Background information and questionnaire (students)

Thank you for being willing to participate in this study on your perceptions as a learner of a foreign language! The following questionnaire should take you between 10 and 15 minutes. If you have any question, please contact the research Dr. Frédérique Grim at Frederique.Grim@colostate.edu.

Name: _____   Age: _____   Gender: M   F

What is/are your major(s)?

_____

What is/are your minor(s), if any?

_____

What year of school are you in?   Circle the appropriate answer(s)

     Freshman   Sophomore   Junior   Senior   Master's
         Doctorate

What is your native language (spoken at home):   English   Other
(Specify): _____

What is the primary foreign language you study: _____

How long (in years) have you studied that specific language: _____

Have you studied/travelled abroad in a country where your primary foreign language was spoken?   Yes   No   If yes, for how long? _____

What class(es) are you currently taking for your primary language?

Question 1: What are the most essential skills to learn a foreign language?

Question 2: What do you need to achieve adequate/successful communication?

Question 3: Rank the following components, in order to importance (1 – most important; 5 least important): Grammar – Communication – Culture – Pronunciation - Vocabulary

Question 4: Do you believe your class(es) prepare you for adequate/successful communication?   Yes   No

     Please, justify your answer.

Notes

[1] **Interpersonal Communication:** Learners interact and negotiate meaning in spoken, signed, or written conversations to share information, reactions, feelings, and opinions.

Interpretive Communication: Learners understand, interpret, and analyze what is heard, read, or viewed on a variety of topics.

Presentational Communication: Learners present information, concepts, and ideas to inform, explain, persuade, and narrate on a variety of topics using appropriate media and adapting to various audiences of listeners, readers, or viewers (The National Standards Collaborative Board, 2015)

[2] ACTFL (American Council on the Teaching of Foreign Languages) and NCATE (National Council for Accreditation of Teacher Education) have set guidelines for the education and preparation of foreign language teachers in the United States, where this research took place.

THE IMPORTANCE OF ASPIRATED INITIAL STOPS
IN ENGLISH AS A LINGUA FRANCA

Mara Haslam, Stockholm University
Elisabeth Zetterholm, Stockholm University

A significant proportion of the population of the world is made up of users of English as a Lingua Franca (ELF). Jenkins (2000) published the Lingua Franca Core (LFC), a syllabus for ELF pronunciation, including the claim that the fortis/lenis distinction must be preserved on English stop consonants for successful ELF intelligibility. The present study evaluates the relationship between Voice Onset Time (VOT) and how the sounds are perceived by ELF listeners. 101 tokens produced during ELF interaction which contained the stops /b/, /p/, /d/, /t/, /g/, or /k/ were played for 9 Swedish listeners, who could indicate that they heard either the word or its minimal-pair counterpart, e.g. *bees* or *peas*. The relationship between VOT and perceived stop was analyzed, with the expectation that longer VOTs would be associated with fortis consonants and shorter VOTs would be associated with lenis consonants. Results followed the predicted pattern for /d/ and /g/ but not for /t/ and /k/. In addition, the pattern observed for /p/ and /b/ is the reverse of the pattern found for the other consonants. These results suggest that further research into the LFC's claim about the fortis/lenis distinction and other LFC claims are warranted.

INTRODUCTION

An estimated one quarter to one third of the world's population speaks English, making it a very important language in the world (Crystal, 2008). Of these speakers, an estimated 75 percent are non-native speakers. Many of these people will engage in communication with other non-native speakers of English, making English as a Lingua Franca (ELF) an important form of communication, worthy of research.

ELF has only recently emerged as a field of study and much remains to be learned about ELF phonology. In 2000, Jenkins published the seminal book "The Phonology of English as an International Language" in which she investigated phonological factors that contribute or do not contribute to misunderstanding in ELF communication. Her conclusions are presented as the Lingua Franca Core (LFC), a list of phonological characteristics which Jenkins claims ELF speech must have in order to be intelligible. The LFC is the only current standard available for ELF pronunciation and has even been included in popular pronunciation textbooks (e.g. Celce-Murcia, Brinton, Goodwin, and Griner, et al., 2010; Hancock, 2012).

However, there is reason to believe that the LFC warrants further testing. First, the data upon which the LFC was based was recorded for other purposes, not specifically for the purpose of investigating intelligibility during ELF communication. The participants in the study upon which the LFC is based were principally Japanese users of English communicating with Swiss German-speaking users of English and therefore represent a somewhat narrow cross-section of ELF users. In addition, Jenkins discovered the phonological points outlined in the LFC by listening for communication breakdown and then attempting to interpret which aspect of the pronunciation caused the breakdown rather than testing the cause of unintelligibility, perception, more directly. While the LFC represents a good step forward in understanding ELF intelligibility and intelligibility of non-native speech in general, it is perhaps prudent to view the LFC as a hypothesis which can be tested, a starting point from which further research can depart. Deterding (2013) uses a methodology similar to Jenkins' (2000), based on external observation of communication breakdown. The present study represents an attempt to apply laboratory methods to one of the aspects of the LFC with the purpose of testing and refining the LFC to further our understanding of ELF phonology and second-language phonology in general. We have turned our attention first to the fortis-lenis distinction on stop consonants as a starting point for investigating the LFC.

One of the claims of the LFC is that ELF intelligibility will break down unless the fortis-lenis distinction on stop consonants is preserved, such as the distinction between /b/ (lenis) and /p/ (fortis). The fortis-lenis distinction is a well-established but somewhat difficult to quantify characteristic of native speakers' speech in English. Lisker and Abramson (1967) identified that for native speakers and for word-initial stops, the fortis-lenis distinction is mostly reflected in Voice Onset Time (VOT), with lenis stops showing shorter VOTs than their fortis counterparts. Therefore the present study takes as its research questions the following:

- Do ELF listeners perceive stops with short VOT as lenis and stops with longer VOTs as fortis?
- Is the fortis-lenis distinction necessary for ELF intelligibility?

METHODS

Stimuli

Word tokens from the Wildcat corpus' (VanEngen, Baese-Berk, Baker, Choi & Bradlow, 2010) Diapix task were used. In this task two participants looked at similar images and talked with each other to find the differences between the images. In order to capture ELF pronunciation, only Diapix conversations in English between speakers with different L1s were used. Table 1 shows the language backgrounds of the talkers and how many talkers of each native language created these conversations.

67

Table 1

First languages and number of talkers from that language background who produced the tokens used in the present study

Native Language	Number of Talkers
Korean	8
Chinese (Mandarin)	6
Spanish	2
Italian	1
Japanese	1
Russian	1
Persian	1
Thai	1
Marathi/Hindi	1
TOTAL	**22**

From these recordings, the monosyllabic words starting with /p/, /t/, /k/, /b/, /d/, and /g/ which were each said by the most talkers were identified and selected as target words. This was done by compiling all words in the corpus from the corpus transcription and then manually selecting the words which fit the requirements. Table 2 shows the target words and the number of talkers who said that target word during their Diapix conversation. The first occurrence of the word in the Diapix conversation for each talker was selected as the token used for the present research. The total number of tokens was 101. No non-target words were used; all items in the experiment were recordings from the corpus. Recordings of the minimal-pair counterparts of the target words (see **Procedure**, below) were not used because such recordings did not always exist in the corpus.

Table 2

Words and number of tokens of each word used in the present study

Word	Number of tokens used
bees	15
cat	20
door	21
got	6
pet	19
two	20
TOTAL TOKENS	**101**

Listeners

Nine listeners were recruited from the Stockholm University community. All identified as native speakers of Swedish. Each listener received a coupon good for a beverage and a pastry at a local café for his/her participation. Swedish-speaking listeners were used in order to simulate an ELF listening situation as much as possible, considering that none of the talkers were native speakers of Swedish.

Procedure

The experiment took place in a quiet office environment. The main task consisted of an identification task where two words in orthographic representation appeared on a computer screen while the listener heard one of the token recordings. Listeners were instructed to press a key labeled "RIGHT" if the word they heard was the word displayed on the right side of the screen and a key labeled "LEFT" if the word they heard was the word displayed on the left side of the screen. The words displayed were always the target word associated with the token and the other member of the minimal-pair containing the voiced counterpart if the word started with a voiceless consonant, or vice versa. For example, if the item involved a token where the target word was "bees", the words "bees" and "peas" were displayed on the computer screen. Each of the 101 tokens was tested in this fashion. The experimental task took approximately 10 minutes to complete.

Table 3 shows the target words and their minimal-pair counterparts. Display of the words was randomized so that the target word did not always appear on the left or right side of the screen. A response was counted as correct if the participant pressed the key corresponding to the side of the screen where the target word was displayed. A response was counted incorrect if the participant did not respond within 4 seconds or if the person pressed the key corresponding to the side of the screen where the minimal-pair counterpart of the target word was displayed.

Table 3

Words used in the study (see Table 2) and their minimal-pair counterparts

Target Word	Minimal-Pair Counterpart
bees	peas
cat	gat
door	tore
got	cot
pet	bet
two	do

After completion of the identification task, the participants filled in a computerized survey with information about their age and background information. Participants were 8 females and 1 male, ranging in age from 21 to 47. All were native speakers of Swedish.

VOT for each of the tokens was measured by one of the authors using the analysis program Praat (Boersma, 2001) and calculations comparing the perception results and the VOT measurements were performed.

RESULTS

Figure 1 depicts, for each target phoneme, the mean proportion of tokens correctly identified as the target phoneme by the listeners.
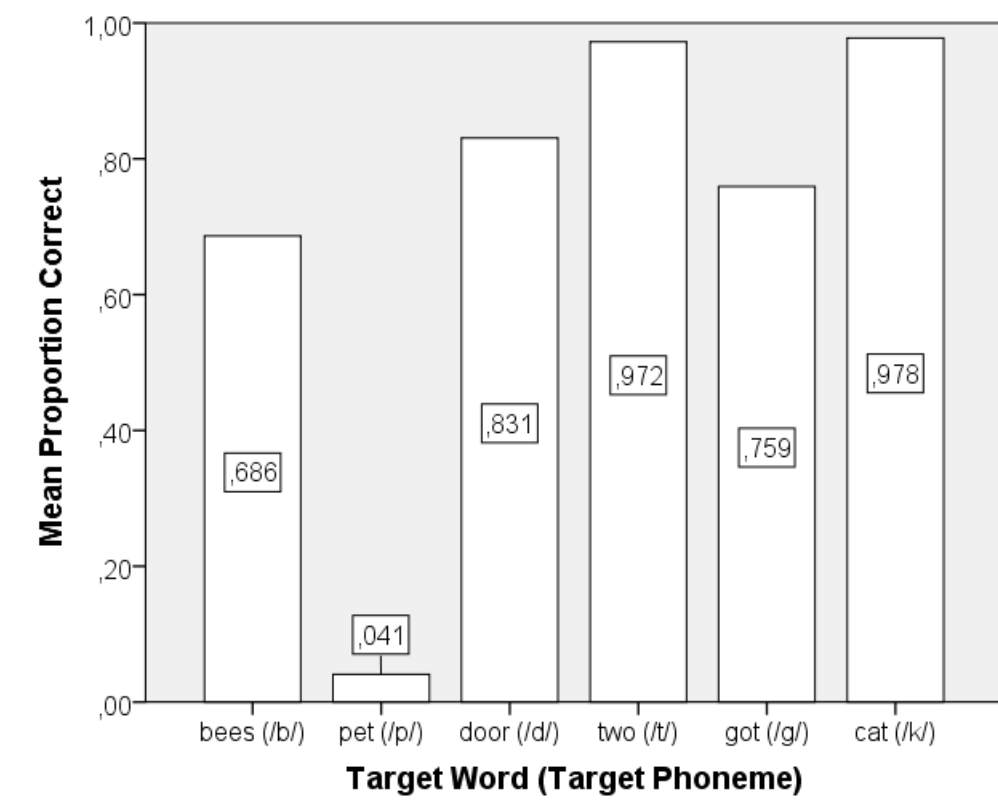
Figure 1. Proportion correct responses (i.e. tokens identified as the target) for the six target phonemes /b/, /p/, /d/, /t/, /g/, and /k/ as measured in 101 total tokens.

Note that in Figure 1 the pattern of responses to the six target phonemes is quite different depending on which phoneme pair the target phoneme belongs to. While the proportion of correct responses to /d/, /t/, /g/, and /k/ were all quite high, the correct responses to /b/ were 69% and the proportion of correct responses to /p/ were 4%, much lower than chance. One–way ANOVA calculation with target phoneme as a within-subjects factor and proportion correct as the dependent variable showed a significant effect of target phoneme (F=106.634, p=0.00) and subsequent comparison of phoneme pairs found significant effects of target phoneme for all three pairs /b/-/p/, /d/-/t/, and /g/-/k/, indicating that participants responded significantly differently to different members of the pairs (/b/-/p/ F=152.407, p=0.00; /d/-/t F=22.708, p=0.001; /g/-/k/ F=17.966, p=0.003).

Figures 2, 3, and 4 depict the relationship between VOT and which phoneme in each pair was selected for /b/-/p/, /d/-/t/, and /g/-/k/ respectively.
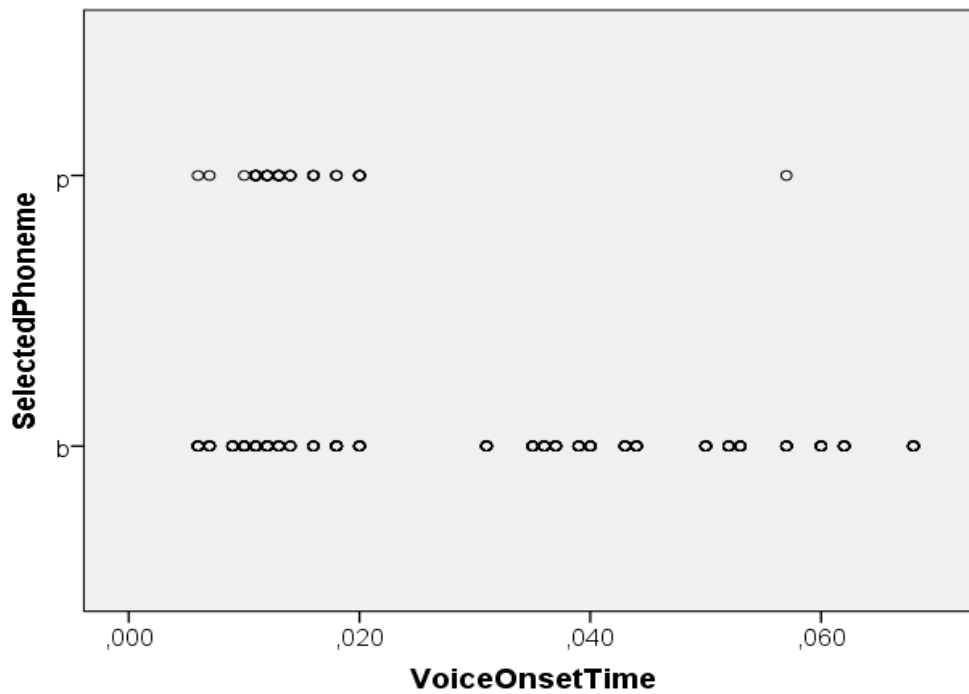
Figure 2. Plot for /b/-/p/ showing VOT and which phoneme in the pair was selected.
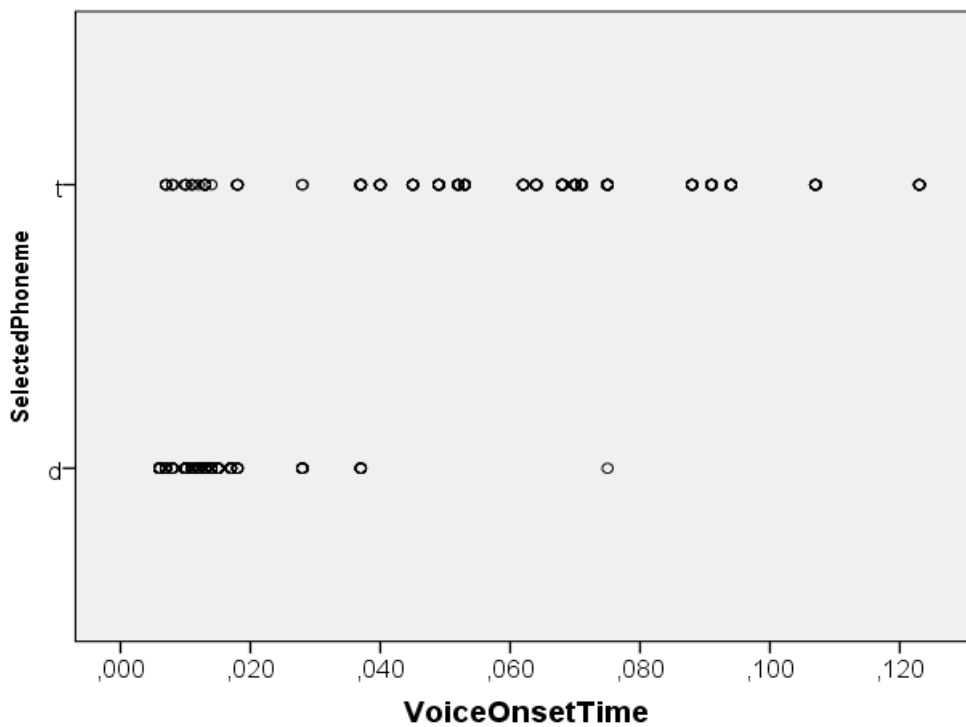
Figure 3. Plot for /d/-/t/ showing VOT and which phoneme in the pair was selected.
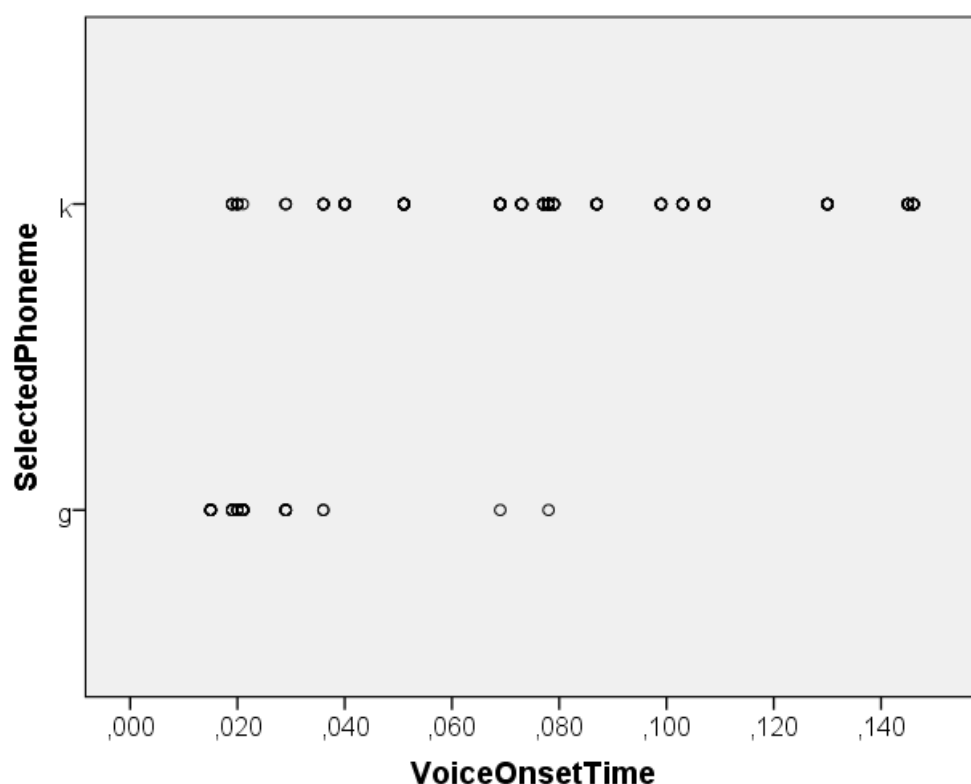
Figure 4. Plot for /g/-/k/ showing VOT and which phoneme in the pair was selected.

The hypothesized pattern was that the lenis member of each pair (i.e., /b/, /d/, or /g/) would be selected for items shorter VOTs and the fortis member of each pair (i.e., /p/, /t/ or /k/) would be selected for items with longer VOTs. For the /d/-/t/ pair and the /g/-/k/ pair this hypothesis is partially followed in that the lenis member of the pair was generally chosen only for short VOTs, but participants selected the fortis member of the pair for both shorter and longer VOTs. However, this pattern is reversed in the /b/-/p/ pair, where /p/ was generally selected only in combination with short VOTs while /b/ was selected for both longer and shorter VOTs.

DISCUSSION

This study was conducted primarily to examine the LFC's claim that the fortis-lenis contrast is important to preserve in ELF speech and that not preserving this contrast would cause problems in intelligibility. The results from this study provide a mixed view on the question, partially supporting the LFC's requirement for the fortis-lenis contrast and partially providing evidence against the LFC's claim. As shown above, listeners were not always successful at identifying whether they were listening to the target word or whether they were listening to its minimal-pair counterpart, especially in the case of /b/-

73

/p/ words. Listeners' relatively poor performance on this pair of sounds would cause them to need to depend on something other than segmental phonology for understanding.

Taking the results into account in general, we can say that participants responded to the /b/-/p/ pair quite differently than they did to the other pairs, and this itself provides important clarification of the LFC. The LFC treats stop consonants as a unified group, arguing that the fortis-lenis distinction is important for all stops, while this study demonstrates that the listeners responded very differently to the /b/-/p/ pair of stops than they did to the other two pairs. This suggests that a more granular approach to the LFC is warranted: instead of treating sounds in large groups such as "stop consonants" or "vowels" perhaps more specific examination will reveal other unexpected information about certain sounds or contrasts that act as intelligibility pitfalls for ELF listeners.

Since the VOT vs. selected phoneme results are noticeably different from the expected pattern of selecting the fortis member of the phoneme pair for long VOTs and the lenis member of the pair for short VOTs, further research will hopefully be able to pin down what listeners are responding to when making their selections. In addition to VOT, a probable candidate is transitions into the following vowel, length of the following vowel, or perhaps some combination of these factors with VOT.

Another issue for further research is the issue of how accurate at perception ELF listeners really are. ELF research in general shows that not much conversational repair is needed and this is taken as evidence that ELF communication is relatively successful. However, the results of this study demonstrate that perhaps ELF listeners are not as successful at perception of consonants as they seem to be. Further research with more controlled methods is warranted to investigate just how successful bottom-up phonological processing in ELF communication is. Perhaps ELF listeners, because of their poor stop consonant perception, are required to depend heavily on top-down processing in ELF communication situations. This has potential implications for both future ELF research and English pronunciation and listening instruction.

ABOUT THE AUTHORS

Mara Haslam is a Senior Lecturer at the Department of Language Education, Stockholm University. Her primary research interest is in second-language phonology with a view to how pronunciation instruction can be improved. This includes interest in second-language intelligibility.

Elisabeth Zetterholm is a Senior Lecturer and Associate Professor in Swedish as a Second Language at Stockholm University. Her current research interests focus on pronunciation in second language learning, comprising segmental and prosodic issues. Zetterholm is also interested in background languages' influence on third language learning, with a focus on phonology transfer and pronunciation. When teaching, most of her students are prospective teachers of Swedish as a second language.

REFERENCES

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5*(9/10), 341-345.

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

Crystal, D. (2008). Two thousand million? *English Today*, *24*, 3-6. doi:10.1017/S0266078408000023.

Deterding, D. (2013). *Misunderstandings in English as a lingua franca: An analysis of ELF interactions in South-East Asia* (Vol. 1). Walter de Gruyter.Hancock, M. (2012). English pronunciation in use intermediate with answers. Cambridge University Press.

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

Lewis, M. Paul, Gary F. Simons, and Charles D. Fennig (eds.). (2015). *Ethnologue: Languagesof the World, Eighteenth edition*. Dallas, Texas: SIL International. Online version:http://www.ethnologue.com.

Lisker, L., & Abramson, A.S. (1967). Some effects of context on voice onset time in English stops. *Language and Speech*, 10, 1-28.

Seidlhofer, B. (2004). 10. Research perspectives on teaching English as a lingua franca. *Annual Review of Applied Linguistics*, *24*, 209-239. doi:10.1017/S0267190504000145.

Van Engen, K. J., Baese-Berk, M., Baker, R. E., Choi, A., Kim, M., & Bradlow, A. R. (2010). The Wildcat corpus of native- and foreign-accented English: Communicative efficiency across conversational dyads with varying language alignment profiles. *Language and Speech*, *53*(Pt 4), 510–540.

THE EFFECT OF METACOGNITIVE STRATEGY INSTRUCTION
ON L2 LEARNER BELIEFS AND LISTENING SKILLS

Jennifer A. Lacroix, Boston University
Marnie Reed, Boston University
Allen G. Harbaugh, Boston University

This pilot study investigated the effect of semester-long strategy-based instruction on learner beliefs and skills in the processing of aural input by adult learners of English as a second language at metacognitive and procedural levels. The study addressed two frequently encountered learner beliefs thought to impede L2 processing of aural input: The little words aren't important; intonation is merely decorative. Working on the premise that learner beliefs underpin learner strategies for processing aural input and are reflected in learner productive and receptive skills, pre- and post-instruction instruments measured both learners' awareness of connected speech processes and the functions of intonation, and their ability to segment a continuous speech stream, and to process utterances for speaker intent. Findings using repeated measures analysis of variance support strategy-based metacognitive training in connected speech and stress and intonation to promote listening skills awareness, aid word segmentation, and facilitate understanding utterance content and intended meaning.

INTRODUCTION

This pilot study investigated the effectiveness of semester-long Strategy-Based Metacognitive Instruction (SBMI) to address learner beliefs and improve listening skills in adult second language (L2) learners of English. As noted by Vandergrift and Goh (2012), learner surveys identify two barriers to effective listening comprehension: word segmentation skills resulting in inability to recognize known words in continuous speech, and failure to grasp message meaning despite understanding the words. Learners who have studied words in isolation in their citation form may not be aware that those words will sound different in connected speech. They may attribute their listening challenges to the speed of speech rather than connected speech features (CPSs), which Alameen and Levis (2015) assert must be addressed in second language teaching. This learner belief - native speakers speak too fast; if they spoke slower I could understand them - may restrain learners from rapidly developing parsing strategies to address one of the true listening challenges: segmenting continuous speech. Although learners may not have reflected on these or similar CPSs in their own language, connected speech processes present similar difficulties for learners of any language in which they occur (Pinker, 1995).

Processing listening input also requires interpreting prosodic elements such as stress and intonation in order to understand a speaker's implied or intended meaning. Defined as "the systematic and linguistically meaningful use of pitch movement at the phrasal or suprasegmental level" (Pickering, 2012, p. 280), intonation presents challenges for L2 leaners. Learners who rely on understanding words for message meaning when processing aural input may not realize that in English, intonation can change the meaning of an utterance and add an implication. Awareness

that English employs even wider pitch ranges to convey emphasis, contrastive or new information, and unspoken but implied information allows listeners to go beyond the literal meaning of an utterance and infer what is meant by what is said. However, learners may be unaware of the power of intonation to "undermine the words spoken" (Wichmann, 2005, p, 229). For example, the fact that the affirmative words "You *can*" (with rise/fall pitch contour) in a teacher's response to the request "Can I turn in my homework late?" are sending a negative message—is generally quite revelatory to L2 learners. Learners who are insensitive to prosodic cues to signal speaker intent instead "may think that intonation is simply decorative" (Gilbert, 2014, p. 125). This learner misconception further highlights the need for a metacognitive approach. Potential benefits of metacognition were recognized by Moyer (2014) who identified a metacognitive approach to language learning as one of five critical factors that account for the success of exceptional learners in her study of second language phonology. This study explores the benefits of a metacognitive approach beyond exceptional populations.

Aim of the Study

The aim of this study was to investigate the efficacy of a metacognitive strategy-based approach to address two challenges to effective listening comprehension: segmenting continuous speech, and processing utterances for speaker intent. Specifically, we investigated whether metacognitive training in connected speech increased learner awareness, strategy use, and skills necessary to aid word segmentation, and whether metacognitive training in contrastive stress and intonation improved learners' accuracy with suprasegmental features, including deducing a speaker's intended meaning.

A pre-instruction needs analysis provided a baseline assessment of learner beliefs and strategies (metacognitive level) and learner skills (procedural level) with respect to word segmentation and interpreting speaker intent. Formative assessments consisting of four written metacognitive reflections and accompanying listening tasks were administered during the instruction phase. The summative assessment, consisting of the final metacognitive reflection and listening tasks, was administered at the end of the course. In order to determine the nature of change, we compared results of the pre-instruction baseline assessment of learner beliefs, reported strategies, and skills to the results of both formative and summative post-instruction assessments of metacognition and skills.

STRATEGY-BASED METACOGNITIVE INSTRUCTION (SBMI)

This study adopted an SBMI instructional approach to teaching listening skills. "Metacognition refers to listener awareness of the cognitive processes involved in comprehension, and the capacity to oversee, regulate, and direct these processes (Vandergrift & Goh, 2012, p. 23). However, in the words of Mendelsohn (2006), "Much of what is traditionally *mis*-named teaching listening should in fact be called testing listening" (p. 75). Thus, the approach adopted in this study employed strategy instruction, the use of meta-language, metacognitive diagnostics and assessments to frame instruction (Reed & Michaud, 2015), and a principled approach to providing corrective feedback in the form of metalinguistic "prompts" (Lyster, Saito & Sato, 2013) to scaffold learning and increase leaners' metacognitive ability to monitor and regulate their progress.

Research Questions

We investigated learners' metacognition and skills with respect to connected speech processes and ability to segment continuous speech, and functions of intonation and ability to process utterances for speaker intent. Our main research question was: What is the nature of change in metacognition and performance before, during, and after the Strategy-Based Metacognitive Instruction (SBMI). To address the aims of the study, the following research questions address learner performance segmenting connected discourse (RQ# 1-3b) and detecting speaker intent (RQ#4).

RQ#1: Will student performance change significantly over time? If confirmed that time is an influential predictor of performance with SBMI:

RQ# 2a: Will there be significant improvement in performance from pre- and post-instruction assessment scores?

RQ #2b: Will student performance on near-immediate assessments show little, if any, significant improvement (thus, change is gradual and not spontaneous).

RQ#2c: Will higher performance persist after the end of the SBMI instructional period (that is to say, will the post-test differ from the mid-intervention measures)?

RQ#3a: Will a moderate to large effect size be observed for time engaged with SBMI.

RQ #3b: Will a moderate to large effect size be observed for increase in performance scores from pre-to post-instruction. Research Question #4: Will a higher percentage of participants respond correctly on the post-instruction assessment?

METHODS

In this study, effects of a metacognitive strategy-based instructional approach on L2 learner beliefs and listening skills were investigated by a pre- and post-assessment comparison.

Participants and Procedures

The study was conducted in a naturally occurring education context under an action research paradigm in one intact elective pronunciation, speaking, and listening elective class taught by the first author. The course was situated within a large, urban, university-based Intensive English Program (IEP) in the United States. The class met semi-weekly for 2.5 hours per session in a 12-week semester.

Subjects

Subjects (n = 14) were in the combined intermediate and low-advanced level section of the course based on their placement scores (56—87 / 100) on the Michigan Test of English Language Proficiency. There were four males; ages ranged from 18-45. Students' national identities were: eight Japanese, two Chinese, one Taiwanese, Korean, Kazakh, and Brazilian.

Instruction

The strategy-based metacognitive instruction was introduced during the first half of the instruction phase, weeks two through six. Practice opportunities were provided; metalinguistic prompts were used to scaffold learning. To facilitate understanding message content, learners were introduced to the following strategy:

Use Three Kinds of Information to Process Aural Input

1) Background (Context) Information
       — what you already know about the topic of conversation

2) Language Information
       — what you know about how the English language works
       (the grammar, vocabulary, and sound system)

3) Acoustic (sound) Information
       — the sounds you actually hear someone saying

Figure 1. Strategy for processing utterance content.

This strategy was represented visually by the instructor, whose corrective prompt consisted of tenting her fingers into a triangle to remind learners to use the appropriate kind of information.
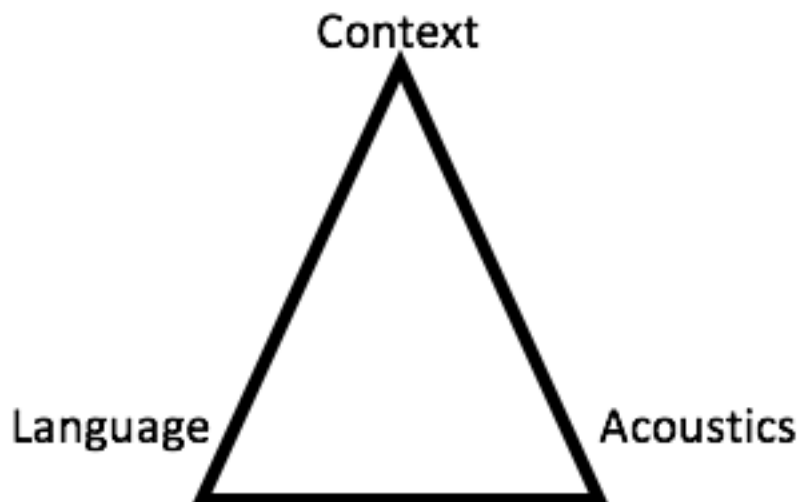

Figure 2. Metalinguistic prompt for using three kinds of information.

To facilitate understanding message meaning, learners were introduced to the following strategy:

Use Three Kinds of Information to Detect Speaker Intent

1) **Detect the Signal**: Marked ("extra") Pitch Range

2) **Locate the Signal**: exaggerated Content or Function word(s)

3) **Interpret the Signal**: Attribute Speaker Intent
    — **Given vs. New Information**
        • Known information; New information
    — **Emphatic Stress**
        • Same Meaning, just emphasized: $X=X^n$
    — **Contrastive Stress**
        • +/- stated, Different Meaning: not X, Y
    — **Implicational Stress** (Implicational Fall-Rise, Wells 2006)
        • Unstated & not retrievable from the locution alone: X+Y

Figure 3. Strategy for processing speaker intent.

Assessment

Pre- and post-instruction metacognitive and skills assessments were administered during weeks one and 12 to determine and monitor learner awareness of CPSs and pragmatic functions of intonation, self-reported strategy use, and performance on listening and speaking skills tasks. Both beliefs and skills were assessed since "learner beliefs affect the range of language learning strategies employed and also affect the motivation to learn, thereby indirectly influencing L2 learning outcomes" (Nix & Tseng, 2014, p. 114). Listening tasks and metacognitive reflections were administered at regular intervals throughout the semester.

Metacognitive Assessment

This study used a combination of measures, including open-ended prompts, true/false and likert-scale judgments, to establish pre- and post-instruction learner beliefs and strategies for processing aural input. The metacognitive reflections prompted speculation about listening challenges and the strategies students used to address them, self-assessment of current strengths and areas for improvement at global and local levels, and identification of aspects about speaking and listening learned in the course.

Skills Assessment

Skills diagnostics targeted ability to segment connected discourse and detect speaker intent conveyed via contrastive stress and intonation. Cloze tasks with accompanying multiple choice and true/false comprehension questions were used to assess connected speech processing. For the

pre-instruction diagnostic, students heard a three-sentence (64-word) prologue to a podcast (Glass, 2010). The instructor provided context information in the form of a brief scripted introduction while the students looked over the cloze passage transcript. Of the 64 words, 46 were supplied; 18 were blanks. Unlike conventional cloze passages, wherein blanks occur at intervals of every nth word, here double blanks occurred, consisting primarily of a preposition and determiner *his*, and four instances of pronoun *he* and auxiliary *was*. This double blank format, characterized by selective replacement of function words, e.g., articles and conjunctions, and the occasional auxiliary or main verb, was used throughout the study. Beginning week seven, students heard portions of a 1-minute 19-second self-contained segment of a TED Talk (Robinson, 2010), divided into four sections of roughly equal duration and word count, with roughly the same number and nature of cloze blanks. Task-specific vocabulary words, phrases, and idiomatic expressions were incorporated into curricular material and practiced extensively throughout the first half of the semester to ensure familiarity at the times of testing. The cloze activity provided an opportunity to gauge the extent to which students were applying the strategies practiced during the first half of the semester for decoding connected speech.

Multiple measures were used to assess learner skills vis-à-vis the use of contrastive stress and intonation to convey speaker intent. Assessments included speaking as well as listening tasks premised upon the assertion by Reed & Michaud (2011) that "the route to successful listening comprehension is through auditory feedback wherein the learner's own increasingly target-like speech production facilitates and reinforces perception" (p. 95). We report here on one such task.

RESULTS

Results are reported for the listening diagnostic task, administered a second time immediately following the 100% failure rate on the first administration: four of the 14 students scored 50% and the others scored between 0% and 44%. Performance on the comprehension questions, answered after the second administration, was at chance or below. The primary data here focus on the cloze tasks results for learners' ability to partially segment speech and on one of the skills assessments for detecting contrastive stress and interpreting speaker intent. Though it should be noted that more data from this pilot study are yet to be analyzed, here we address the main research questions. To address RQ1, a repeated-measures (single within-subjects factor) ANOVA was conducted to evaluate changes in cloze task scores over time. The within-subjects factor was the testing administration times (one pre-instruction, one post-instruction, and 4 mid-instruction assessments). The dependent variable was the percentage of words correct on a cloze activity (ranging from 14 to 20 items per assessment). Comparable results were obtained running the RM ANOVA with percentages as with arcsine transformations; for ease of presentation, the results reported here are for the untransformed percentages.

 There was missing data for three of the n=14 students, and there was a single potential univariate outlier in the pre-instruction assessment. Missing data were imputed (see Appendix for details on imputation protocol), and results were run with and without the imputed data; subjects were deleted list-wise if missing data. Comparable results were obtained in all but one instance; results reported here are for the imputed data set. Regarding the outlier, results were only more strongly supported with the removal of the outlier, and consequently it was retained in all analyses reported here. All other model testing assumptions were confirmed for this data set.

The repeated measures ANOVA indicated a statistically significant within-subjects effect for time with $F(5,65)=8.743$, p<0.001; $\omega^2 = 0.228$ suggests this to be a moderate effect. Summary statistics for the time points were $M_{pre} = .34$ (SD=0.27), $M_1 = .63$ (0.13), $M_2 = .58$ (0.17), $M_3 = .62$ (0.24), $M_4 = .61$ (0.19), and $M_{post} = .74$ (0.17). Aggregated results are presented in Figure 4 along with select student exemplars (chosen for low, 30th percentile, 70th percentile and high performance).
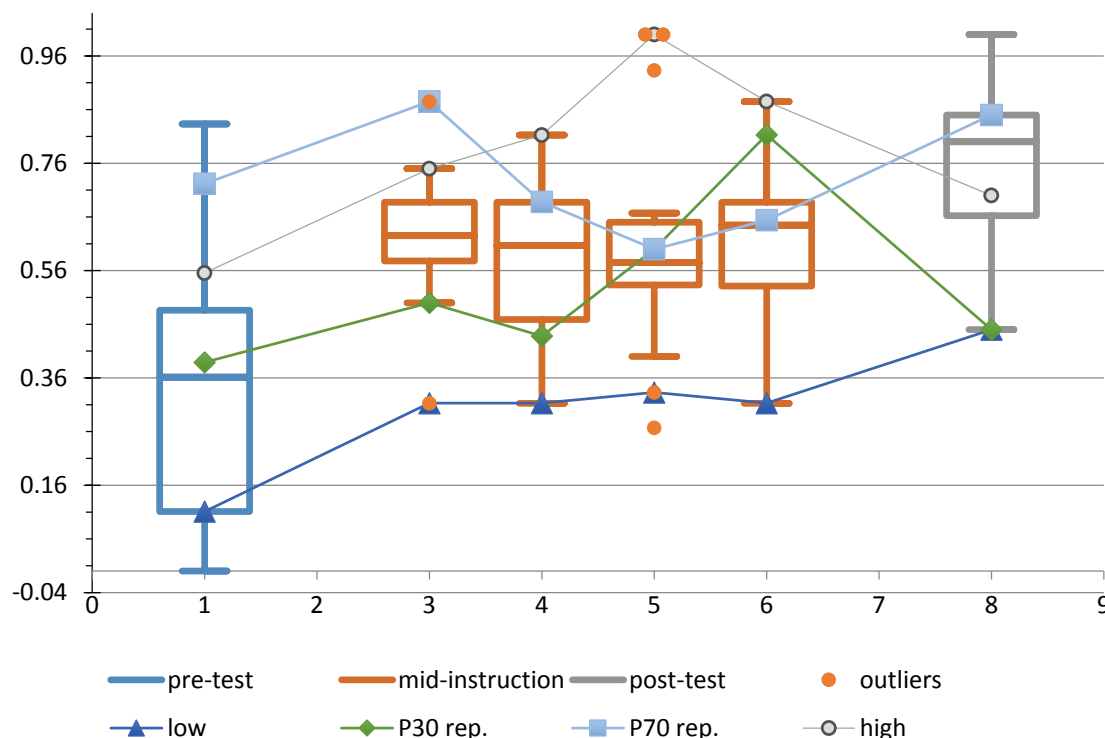
Figure 4. Student performance (as percentage correct) for cloze tasks over time.

Following on from this (omnibus) finding, select post hoc analyses were conducted to confirm which time points differed from others. First, it was confirmed there was indeed a difference between the pre- and post-assessments with $t(13) = 5.085$, $p < .001$; the effect size was

$d = 1.36$ ($M_\Delta = 0.40$, $SD_\Delta = 0.30$). Next, it was observed there was no statistically significant difference across the four assessments that occurred during the instructional phase with $F(3,39) = 0.281$, $p = .839$. A statistically significant finding was observed when comparing pre- and mid-instructional assessments, $F(4,52) = 6.849$, $p < .001$. Additionally, a statistically significant finding was observed when comparing the mid- and post-instruction, $F(4,52) = 2.720$, $p = .039$. (However, it should be noted that this finding was not corroborated when using only non-imputed data, $p = .32$.)

To address RQ 4, a Fisher's exact test, conducted due to the small sample size, found a significant difference between pre- and post-instruction assessments ($p < .001$) on use of prosodic cues to determine speaker intent. Following a pre-test using cardinal numbers, students

practiced throughout the semester producing differential stress and intonation when providing information versus clarifying misinformation (e.g., on tasks involving numbers). In the post-test, students were asked to judge whether a speaker is clarifying misinformation about a zipcode. Answering in the affirmative suggests detection of contrastive stress to signal clarification.

Table 1

Student Responses for the Information Clarification Task

	Pre-instruction	Post-instruction	Total
Correct responses	1	10	11
Incorrect responses	9	1	10
Total	10	11	21
N/A	4	3	7

The proportion of students answering in the affirmative post-instruction (10 of 11) is higher than the proportion answering in the affirmative pre-instruction (1 of 10), z = 2.45, p = .007. Contextually, of the 8 students providing both pre- and post-responses, 6 of the 7 indicating no at the pre-test reported yes at the post-test.

METACOGNITION SUMMARY

Regarding CPSs, learners' initial beliefs implicated speed of speech as a primary cause of listening difficulty parsing continuous speech. Thirteen (of 14) subjects agreed that if native speakers spoke more slowly, they could understand utterance content. Diagnostics also revealed a listening strategy, articulated by all subjects: pay attention to content words; little words are not important. This strategy was later revealed to reflect advice from previous instructors in integrated skills and elective classes.

Regarding the communicative and pragmatic functions of intonation, a seven-point true/false questionnaire was used to assess learners' pre- and post-instruction beliefs. As evident from Table 3, most students initially agreed upon two of the functions, but the number in agreement decreased as the complexity of the functions increased.

Table 2

Results of Pre-Instruction Metacognitive Assessment of the Functions of Intonation

	True	False	**Prompt:** In English, intonation can . . .
1.	13 (93%)	1	turn a statement into a question
2.	13 (93%)	1	act as oral punctuation, quotation marks, and paragraph breaks
3.	10 (71%)	4	turn a sincere statement into a sarcastic one
4.	10 (71%)	4	signal an implied contrast
5.	9 (64%)	5	change the meaning of a sentence
6.	9 (64%)	5	reduce the number of words needed to convey your meaning
7.	8 (57%)	6	convey information without actually saying the words

In this analysis, we mainly focused on pre- and post-instruction assessments. Below are data for the metacognitive assessment of the pragmatic functions of intonation.

Table 3

Results of Post-Instruction Metacognitive Assessment of the Functions of Intonation

	Mean	SD
Day 1 Metacognition T/F	77.92	20.72
Final Metacognition T/F	84.64	19.04

While there was no significant difference between the pre- and post-test measure, $t(12) = 1.20$, $p = .25$, it was noted that the mean of the students' scores on the metacognition T/F measure did increase and the effect size was $d = 0.32$. Given a small to moderate effect size such as this, it is possible that the reason there was no statistically significant difference was due to the small number of students in this sample.

DISCUSSION

Preliminary findings of this study suggest that strategy-based metacognitive instruction resulted in improved listening skills awareness, aided word segmentation in continuous speech to facilitate understanding utterance content, and aided detection of contrastive stress and intonation to facilitate understanding message meaning.

The first research question regarding change in student performance over time was affirmed ($p < .001$). Following from this, it was also confirmed that there was a significant improvement in performance from pre- and post-instruction assessment ($t(13) = 5.085, p < .001, d = 1.36$). As

was expected regarding research question #2b, while it was confirmed that time is an influential predictor of performance with SBMI, student performance on near-immediate assessments showed no statistically significant improvement with regard to time ($p = .84$). Though such a claim cannot be proven, it is noteworthy to suggest that change during instruction was small and incremental, but substantive in the duration. To support this, it was observed via the repeated measures ANOVA that the pre-instruction scores were noticeably different from the mid-instruction scores ($p < .001$). Specifically, addressing research question # 2c, will higher performance persist after the end of the SBMI instructional period, the results were inconclusive ($p = .039$ w/imputed data and $p = .32$ with list-wise deletion of missing data). However, an examination of the data clearly suggests that performance did not worsen after instruction ended. With respect to research questions, #3a & # 3b, a moderate to large effect size, ($\omega^2 = 0.23$) was observed for time engaged with SBMI, and a large effect size ($d = 1.33$) for increase in performance scores from pre- to post-instruction was observed. Addressing research question #4, a change in student performance interpreting speaker intent was observed, with a significantly higher number of participants responding accurately at the post-test.

Qualitative Findings

Though not reported here, qualitative data obtained from learners' post-instruction metacognitive reflection indicated robust awareness of the importance of connected speech processes and stress and intonation along with an articulated use of metalinguistic knowledge closely paraphrasing the SBMI language.

Limitations of the Study

The study consisted of a convenience sample of students enrolled in one section of a pronunciation, speaking, and listening elective course. There was no control group.

CONCLUSION

Learners' mean post-instruction metacognitive awareness of connected speech processes was higher than pre-instruction, and in addition to increased metacognition, listening segmentation skills improved as well. Learners' metacognitive awareness of the pragmatic functions of intonation increased; however, due to the small sample size, while the mean scores increased, the increase was not statistically significant. Although skills data are still under analysis, regarding the information clarification task, a statistically significant difference between pre- and post-tests

was found. The encouraging findings of this study support continued investigation of the efficacy of a strategy-based metacognitive approach.

ACKNOWLEDGMENTS

The authors wish to thank Dr. Alan Broomhead for his support, and Di Liu for his assistance with data analysis.

ABOUT THE AUTHORS

Jennifer A. Lacroix is a doctoral fellow of education at Boston University in Developmental Studies (language education). She has over sixteen years of ESL teaching experience, including in IEP settings, and has presented widely at international conferences.

Marnie Reed is Associate Professor of Education and affiliated faculty in the Program in Applied Linguistics at Boston University. Dr. Reed's research interests are applied phonology, particularly the role of metacognition in cross-linguistic awareness of the functions of intonation. She is co-author (with Christina Michaud) of *Sound Concepts: An Integrated Pronunciation Course* (McGraw-Hill), and co-editor (with John Levis) and chapter contributor to the Wiley-Blackwell *Handbook of English Pronunciation*.

Allen G. Harbaugh is a Clinical Assistant Professor of Education at Boston University in Educational Leadership & Policy. He previously worked as a community college mathematics instructor for over 15 years and served as the Academic Chair of Tertiary & Adult Education at Murdoch University in Perth, Australia. Dr. Harbaugh is an international scholar who collaborates with many researchers in a variety of disciplines as a statistical analyst.

REFERENCES

Alameen, G. & Levis, J. (2015). Connected speech. In M. Reed and J. Levis (Eds.), *The Handbook of English Pronunciation* (pp. 159-174). West Sussex, UK: John Wiley & Sons, Inc.

Gilbert, J. (2014). Myth 4: Intonation is hard to teach. In L. Grant (Ed.) *Pronunciation Myths: Applying Second Language Research to Classroom Teaching* (pp. 107-136). Ann Arbor, MI: University of Michigan Press.

Glass, I. (2010, February 19). *Episode 401: Parent Trap, This American Life*. Retrieved from thisamericanlife.org

Lyster, R., Saito, K., & Sato, M. (2013). Oral corrective feedback in second language classrooms. *Language Teaching, 46*(1), 1-40.

Mendelsohn, D. (2006). Learning how to listen using learning strategies. In P. Gorden (Ed.), *Current Trends in the Development and Teaching of the Four Language Skills* (pp. 75-89). Berlin: Mouton deGruyter

Moyer, A. (2014). Exceptional outcomes in L2 phonology: The critical factors of learner engagement and self-regulation. *Applied Linguistics, 35*(4), 418-440.

Nix, J-M. L, & Tseng, W-T. (2014). Towards the measurement of EFL Listening Beliefs with item response theory methods. *The International Journal of Listening, 28*, 112-130.

Pickering, L. (2012). Intonation. In: K. Malmkjaer (Ed.), *The Routledge Linguistics Encyclopedia*, 3rd edition, (pp. 280–286). NY: Routledge.

Pinker, S. (1995). *The Language Instinct: How the Mind Creates Language*, New York: Harper Perennial.

Reed, M. and Michaud, C. (2011). An integrated approach to pronunciation: Listening comprehension and intelligibility in theory and practice. In J. Levis & K. LeVelle (Eds.). *Proceedings of the 2nd Pronunciation in Second Language Learning and Teaching Conference,* (pp. 95-104) Ames, IA: Iowa State University.

Reed, M. and Michaud, C. (2015). Intonation in research and practice: The importance of metacognition. In M. Reed & J. Levis (Eds.) *The Handbook of English Pronunciation* (pp. 454-470). West Sussex, UK: John Wiley & Sons, Inc.

Robinson, K. (2010). Bring on the learning revolution. *TED2010.* Retrieved from https://www.ted.com/talks/sir_ken_robinson_bring_on_the_revolution

Vandergrift, L., & Goh, C. (2012) *Teaching and Learning Second Language Listening: Metacognition in Action.* NY: Routledge.

Wichmann, A. (2005). Please – from courtesy to appeal: the role of intonation in the expression of attitudinal meaning. *English Language and Linguistics, 9*(2), 229–253.

APPENDIX

Missing data were imputed in two separate ways. Missing data from the mid-instruction assessments were imputed using a geometrically weighted average for proximal times (e.g., information from time i±1 was weighted 1.00; time i±2 was weighted 0.50; time i±3 was weighted 0.25; etc.). Missing data from the pre-instruction period was imputed using a multiple regression equation with dependent variables consisting of the post-instruction score and the average of the mid-instruction scores.

IS PHONEMIC TRAINING USING NONSENSE OR REAL WORDS MORE EFFECTIVE?

Ron I. Thomson, Brock University
Tracey M. Derwing, University of Alberta & Simon Fraser University.

In this exploratory study we investigate whether focusing learners' attention on phonetic form through the presentation of vowels in isolated open syllables leads to greater learning than presentation of vowels in real words. Thirty-one intermediate, mixed-L1, English learners were assigned to three experimental conditions: A phonetically-oriented group (n=9); a word-focused group (n=12) and a control group (n=10). Using a high variability phonetic training (HVPT) paradigm, learners were taught to identify ten English vowels. Randomized recordings of the learners' pronunciation before and after training were evaluated by expert judges. Results indicated that the phonetically-oriented training was superior in promoting improvement in the pronunciation of real words relative to training almost entirely focused on the pronunciation of those same words. Furthermore, we found evidence to suggest that the accuracy of pronunciation before and after training was affected by interactions between the lexical frequency of the word in which the vowel appeared and each vowel's surrounding phonetic environment. Finally, implications for teaching are briefly addressed.

INTRODUCTION

The past decade has witnessed a notable increase in second language (L2) pronunciation research. Furthermore, many recent studies examining the efficacy of pronunciation instruction for adult learners indicate that instruction often makes a positive difference (Lee, Jang & Plonsky, 2015). However, although increased attention to what was previously an underrepresented area of inquiry is heartening, many gaps in the breadth and quality of the emerging research remain (Thomson & Derwing, 2015). Specifically, much of the current literature focuses on the efficacy of traditional approaches to teaching pronunciation, or repeatedly investigates the same questions rather than seeking new knowledge. For instance, the number of studies investigating English /l/-/r/ acquisition by Japanese speakers are added to regularly, but with little new insight gained (Thomson, 2011). Thomson and Derwing (2015) argue that the means of assessing development in pronunciation studies should be addressed. They particularly critique the over-reliance on reading tasks to test pronunciation. While easy to administer, these tasks do not provide an ideal representation of learners' speech in the context of real world communication. In this exploratory study, we attempt to fill gaps in the types of questions asked, and in the methods used to assess performance.

Background

It is widely accepted that most adult L2 learners do not develop native-like pronunciation of L2 vowels and consonants (Abrahamsson & Hyltenstam, 2009; Baker & Trofimovich, 2006; Bongaerts, van Summeren, Planken, and Schils, 1997; Munro & Derwing, 2008; Munro, Derwing, & Thomson, 2015). This reality has led many researchers and practitioners to search for the holy grail of pronunciation instruction: a method that can overcome this seemingly insurmountable obstacle. We do not view this search as productive, but instead agree with Levis (2005), who argues that approaches that treat accent-free L2 speech as the goal are both unrealistic and unnecessary. When it comes to the teaching of segmentals, our objective is to help learners better recognize and produce L2 speech sounds in ways that allow variation of pronunciation within individual categories. This approach accepts that how any given speech sound is produced can vary substantially without being misperceived by interlocutors as a member of a different category. Promoting within-category variation is in keeping with the fact that native speakers also differ in how they pronounce the same phonemes in different words or phonetic environments – what phonologists term allophonic variation. One obvious example is the difference in how English /l/ is pronounced in the word 'like' compared to the word 'ball'. Such allophonic variation is not always as obvious, however. In fact, all speech sounds are influenced by their neighboring phonetic environments. Consider the pronunciation of /u/ in 'boo' and /u/ in 'goo.' In the former word the vowel is produced much further forward in the vocal tract than in the latter.

Accepting that within-category variation is natural suggests that teaching the pronunciation of L2 sounds should incorporate and emphasize variation rather than focusing on elusive prototypes, citation forms, and the pronunciation of sounds in isolation. One technique that allows for a controlled approach to input variability is High Variability Phonetic Training (HVPT). Based on laboratory studies by Logan, Lively & Pisoni (1993), HVPT trains learners to identify L2 sounds in the context of stimuli spoken by multiple talkers and in multiple phonetic environments. Learners are first asked to indicate which speech sounds they perceive and are then provided with feedback on the accuracy of their responses. Manipulating training stimuli in terms of the number of talkers or the number of phonetic contexts in which sounds are presented is simple, and allows for determining whether training extends to new talkers and contexts (see Thomson, 2011; 2012a).

Since HVPT provides learners with feedback on the accuracy of their perceptions, it can help to direct their attention to properties of segmental stimuli important for L2 category formation. This is valuable because noticing linguistic forms in natural speech is often challenging for adult learners, who tend to focus on meaning instead of form (Schmidt, 2001). Guion and Pederson (2007) conducted an experiment in which adult L2 learners were trained to perceive L2 Hindi contrasts using identical stimuli, but one group was tasked with learning the meaning of words that differed by a single sound, while another group was asked to pay attention to the sounds of stimuli and how they differed, without knowing their meaning. The sound-oriented group performed significantly better on a discrimination test after training than the group that had been asked to focus on learning the meanings of the words.

The present study extends previous research by examining whether phonetically-oriented HVPT

89

training or lexically-oriented HVPT training has a greater impact on L2 learners' pronunciation of English vowels.

Research Questions

The following three research questions guided our study:

1. Does perceptual training using nonsense words *or* training predominantly focused on real words result in better pronunciation of real words?

2. Does degree of pronunciation improvement depend on the type of assessment task used?

3. Does pronunciation accuracy differ across English vowel categories and phonetic/lexical contexts?

METHODS

Participants

Thirty-six adult immigrants studying fulltime in a Canadian English as a Second Language (ESL) program volunteered. All were selected on the basis of having similarly assessed English abilities (LINC 5 – which is intermediate proficiency). Thirteen were randomly assigned to each of two experimental groups, and ten to a control group. Immediately after commencing the study, four participants withdrew from the first experimental group and one from the other, citing a lack of time to devote to the training portion. This left 31 participants, whose details are provided in Table 1.

Table 1.
Demographics of English learner participants

	Phonetic Group **(n=9)**	**Real Word Group** **(n=12)**	**Control Group** **(n=10)**
Age	M=37.3 (range: 24-46)	M=36.4 (range: 25-48)	M=33.1 (range: 23-40)
Sex	7 female; 2 male	9 female: 3 male	7 female; 3 male
Length of Residence	M= 12.7 months (range: 5-27)	M=20 months (range = 4-96)	M=15.5 months (range 5-57)
L1	Chinese (6); Amharic (2); Spanish (1)	Chinese (4); Spanish (2); Arabic (1); Russian (1); French (1); Romanian (1); Kinyarwandan (1); Punjabi (1)	Chinese (1); Spanish (2); Amharic (1); Arabic (2); Russian (1); French (1); Igbo (1); Zo (1)

On the surface it appears that the three groups are different in terms of their mean Length of Residence (LOR). In fact, these differences are largely attributable to a single outlier in the Real Word and Control groups, as reflected in the LOR range. When the single most extreme LOR outliers are removed from each group, their mean LORs are 13.6 and 10.9 months respectively. It should also be noted that the LOR outliers reported having very little interaction in English outside of the classroom. This, combined with their similar proficiency level, provides some assurance that LOR will not be an important factor in this study.

Perceptual Training

Using English Accent Coach (EAC) (Thomson, 2012b), learners in both experimental groups were trained to better recognize ten English vowels /i, ɪ, e, ɛ, æ, ɑ, ʌ, o, ʊ, u/. This freely available web-based HVPT application (www.englishaccentcoach.com) can be used to present to learners with isolated open syllables (i.e., just a consonant + vowel) or words containing target vowels and consonants, spoken by thirty speakers of Canadian English (similar to General American). Learners must respond to each item by clicking on the phonetic symbol associated with the vowel or consonant they believe each item contains (e.g., the identity of the initial consonant, stressed vowel, final consonant, etc.). After making their choice, they receive auditory and visual feedback on the accuracy of their selection. In this study, a researcher mode of EAC was used to precisely control the stimuli presented to each of the two experimental groups.

The first experimental group (Phonetic Group) received perceptual training for English vowels in the context of isolated open syllables (e.g., /bi/, /pi/, bɪ/ and /pɪ/); only 25% of their training sessions incorporated 70 target words (seven words containing each of the ten vowels). In the case of isolated CV syllables, many resulting tokens are not real words (e.g., /hɪ/, /hɛ/, /hʊ/), thus forcing learners to attend to phonetic information to successfully identify the vowels.

The second experimental group (Real Word Group) was trained almost entirely using the 70 target words. This group received only three brief phonetically-oriented sessions at the outset to ensure that they had learned the phonetic symbols, and a single phonetic session at the end. In CVC or more complex real words, learners are often able to recognize the word, but having recognized it, they may then apply knowledge concerning which vowel is supposed to occur in that word, as opposed to focusing on its phonetic properties.

Participants in both the Phonetic and Real Word groups completed 40 training sessions, at their leisure, over the course of one month, but were told they could complete a maximum of two sessions on any given day. Training sessions 1-3 comprised 100 items each, sessions 4-39, 150 items, and session 40, 200 items. The Control Group received no perceptual training. None of the groups received explicit articulatory training or practice. Like Thomson (2011, 2012a), we assume that improvement in perception will lead to changes in productions without any explicit pronunciation practice. Details concerning the training sessions are provided in Table 2.

Table 2

Perceptual training sessions by group and stimuli

Session	Phonetic Group	Real Word Group
1 – 3	Phonetic training: h+V syllables	Phonetic training: h+V syllables
4 – 29	Phonetic training: Open syllables presented in sets based on related consonant onsets. e.g., Session 1: p+V and b+V; Session 2: g+V and k+V, etc.	Target word training
30 – 39	Target word training	
40	Phonetic training: h+V syllables	Phonetic training: h+V syllables

Production Recordings

Before and after training, two tasks were used to elicit participants' productions of the same 70 target words used in training. In the first, participants heard the target words embedded in the carrier phrase, "The next word is ___," and they responded by repeating the word they had just heard in the carrier phrase, "Now I say ___." In the second task, a twenty-word subset of the target words (two nouns for each of the ten target vowels) were presented in the form of pictures. The participants created a sentence using each word. All tests were administered to participants individually in a quiet room and recorded using a high quality digital recorder and microphone. Individual productions of each target word from both the elicited imitation and picture-naming tasks were later extracted from the long recordings and saved as 5573 separate sound files representing individual productions of individual words (seven tokens were lost due to recording errors).

Judgments of Vowel Intelligibility

The 5573 individual sound files comprising participants' productions before and after training were presented to two phonetically trained judges (the authors) for evaluation using Praat's (Boersma & Weenink, 2016) Multiple Forced Choice script (see Thomson, 2013 for a detailed description). Files were blocked by each of the ten target English vowels, but randomly across speakers, time, and speaking task. These blocks of approximately 550 words each were further subdivided into five sets so that the task was more manageable and could be spread over a number

of sessions and days to avoid fatigue. Judges were asked to respond to each item by indicating whether it was a 'good' or 'poor' exemplar of the intended category, or another category altogether, assigning values of 2, 1, and 0 respectively.

RESULTS

We first examined the extent of agreement between judges. In 71% of cases, both were in complete agreement across the three possible categories (i.e., 'good', 'poor' or 'other'). When the good and poor examples were collapsed into a single category, agreement was at 84%. Given the borderline nature of some productions, where it was practically a coin toss between a very poor rendition of a particular vowel, or a poor rendition of the neighbouring category, this degree of rater agreement is extremely good. Because the judges' responses showed strong agreement, they were averaged for each item.

Three Bonferroni-adjusted Wilcoxon Signed Rank Tests were conducted to measure changes in performance over time on the elicited imitation task for the three groups. These indicated a statistically significant improvement in English vowel pronunciation for the phonetically-trained group ($z = -2.695$, $p = .007$) with a small effect size ($r = .08$). No significant improvement was detected for either the lexically-focused group or the control group.
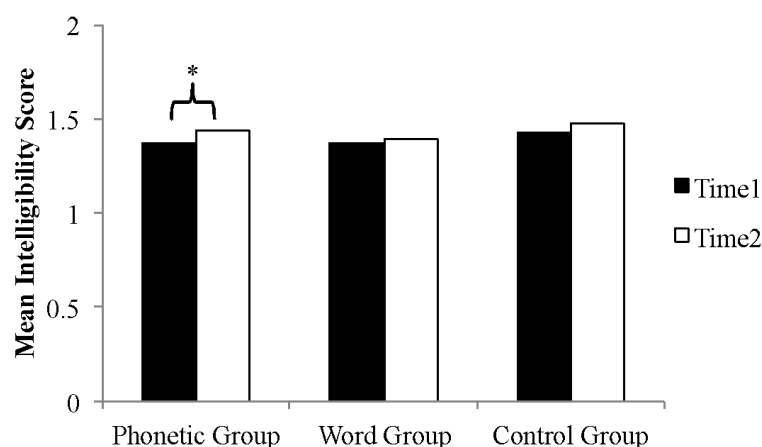
Figure 1. Mean intelligibility scores over time x group. The asterisk indicates a significant difference.

Three Bonferroni adjusted Wilcoxon Signed Rank Tests were also conducted to measure the groups' changes in performance over time on the picture naming production task (i.e., learners used target words in their own sentences). These tests indicated no significant difference for any group, although descriptively, the phonetic group demonstrated a larger improvement in the mean than the other two groups.

We conducted post-hoc analyses to examine whether improvement in the elicited imitation task was limited to particular vowels, or extended across all vowel categories. Learners in the

Phonetic Group demonstrated mean improvement in 8/10 vowel categories, /i, ɪ, e, ɛ, æ, o, ʊ, u/, with means scores for the remaining two, /ɑ/ and /ʌ/, decreasing between Time 1 and Time 2. Learners in the Real Word Group demonstrated mean improvement in 4/10 vowel categories /e, æ, ɑ, o/), with mean scores for the remaining vowels staying unchanged or decreasing. Finally, learners in the Control Group demonstrated mean improvement in 6/10 vowels /ɪ, e, ɛ, æ, ɑ, o/, with mean scores for the remaining vowels each decreasing. In sum, while scores on /i, ʊ, u/ increased for the Phonetic Group, they did not improve for the Real Word and Control Groups. Conversely, the mean score for /ɑ/ decreased for the Phonetic Group, while improving for the other groups.

We next examined individual trajectories in each group, finding that 89% of the Phonetic Group improved over time, with only one participant showing no improvement, but also no decline. In contrast, only 50% of the Real Word Group and 60% of the Control Group demonstrated improvement in their mean scores. There was no correlation between LOR and extent of improvement.

Finally, we examined whether particular L2 vowel categories were more intelligible than others, and whether the word in which they occurred played a role. Recall that each vowel occurred in seven distinct lexical contexts. Results indicated that four vowels /i, e, ɑ, o/ were more accurately produced than the remaining vowels, and that their scores were also least affected by the words in which they occurred. For example, the /i/ in 'bead' was as intelligible as the same vowel in 'leaf'. In contrast, the intelligibility of /ɪ, ɛ, æ, ʌ, ʊ, u/ varied dramatically, depending on the words in which they were produced. For example the /ɪ/ in 'bid' was rarely intelligible, while the same vowel in 'which' and 'stick' was usually intelligible. Similarly, the /ɛ/ in 'head' and 'jet' were very intelligible, while in 'bread' it was not. While word familiarity and frequency may play a role (Thomson & Isaacs, 2009; Munro & Derwing 2008) this does not easily account for much of the variation in the current data. For instance, while the vowel in 'bread' (high frequency) had a low score both before and after training, the same vowel in 'sketch' (low frequency) had a high score at both times. Furthermore, there is no clear evidence that the phonetic environment plays a primary role. For example, the vowel in 'bud' was far less intelligible than the same vowel in 'scum.' The latter has a complex onset, which does not appear to cause difficulty for speakers in producing the following vowel. Similarly, the vowels in 'spin' and 'stick' were both far more intelligible than in the word 'bid'.

DISCUSSION

This exploratory study provides preliminary evidence in response to our three research questions. First, does perceptual training using nonsense words *or* training predominantly focused on real words result in better pronunciation of real words? The findings suggest that at least for participants in this study, forcing learners to attend to phonetic details during perceptual training resulted in significant improvement in pronunciation. However, in answer to our second question, regarding whether improvement depends on the type of assessment task used, we found that detectible improvement is limited to more controlled productions. That is, it did not seem to transfer to more extemporaneous pronunciation of the same words. It is encouraging to see improvement in the elicited imitation task, however, since this task is more challenging than the reading tasks commonly used in this line of research.

94

Finally, we asked whether pronunciation accuracy differs across English vowel categories and across phonetic/lexical contexts. We found that, in general, English lax vowels, /ɪ, ɛ, æ, ʌ, ʊ/ as well as /u/ were the most challenging for learners, while the remaining four tense vowels /i, e, ɑ, o/ were relatively clear, perhaps because similar vowels appear in the learners' L1s. This seemed to be the case regardless of the phonetic or lexical context in which the vowels were found. While there may be some evidence that lexical frequency affects the intelligibility of vowel production, it is not a straightforward predictor. Furthermore, there is little evidence to support a general pattern with respect to vowel intelligibility being affected by the complexity of the surrounding phonetic environment. Therefore, it is more reasonable to conclude that a complex interplay among lexical frequency, phonetic context, and other factors accounts for differences in intelligibility scores for the same vowel category produced in different words. In some cases, a vowel found in a complex phonetic environment may be in a word that is so frequent that the learner had already acquired the vowel in that word.

IMPLICATIONS

Our findings suggest that instruction of L2 vowels should include some focus on phonetic level information, as opposed to focusing solely on the pronunciation of sounds in real words. This approach may help to draw learners' attention to phonetic information in a way that using only words as training stimuli does not. At the same time, the improvement demonstrated by the Phonetic Group in this study had a small effect size, while the Real Word Group showed no improvement after a significant amount of input. These marginal gains suggest that perceptual training on its own is insufficient to promote maximal improvement. Instead, learners likely need explicit practice producing the sounds they are learning to more accurately perceive. While the present study lasted a month, the hours of training was relatively small (approximately 10), and even smaller on a per vowel basis (1 hour). Thus, we speculate that longer training will have a stronger impact.

Finally, this study has implications for focus of training. As Munro and Derwing (2008) and Munro, Derwing & Thomson (2015) found, some English sounds appear to be easy for learners, in some instances, because there are direct parallels in their L1, and in other cases, they may simply be easy to perceive and produce. Knowledge of which categories are most challenging can allow teachers to focus on those sounds that are less amenable to natural improvement.

ACKNOWLEDGEMENTS

ABOUT THE AUTHORS

Ron Thomson is an Associate Professor of Applied Linguistics at Brock University, in Niagara, Canada. He teaches in the undergraduate program in Applied Linguistics, and in the MA TESL

program. His research interests include oral fluency and pronunciation development, oral fluency instruction, high variability phonetic training, and computer assisted pronunciation teaching.

Tracey Derwing is an Emeritus Professor in the TESL program at the University of Alberta, and an Adjunct Professor in the Linguistics Department at Simon Fraser University.   She is interested in identifying ways to enhance communication across L1 and L2 learners.

REFERENCES

Abrahamsson, N., & Hyltenstam, K. (2009). Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language Learning*, *59*, 249–306.

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

Boersma, Paul & Weenink, David (2016). *Praat: doing phonetics by computer* [Computer program]. Version 6.0.14, retrieved 11 February 2016 from http://www.praat.org/

Bongaerts, T., van Summeren, C., Planken, B., & Schils, E. (1997). Age and ultimate attainment in the pronunciation of a foreign language. *Studies in Second Language Acquisition, 19*, 447-465.

Guion, S. G., & Pederson, E. (2007). Investigating the role of attention in phonetic learning. In O.-S. Bohn & M. J. Munro (Eds.), *Second-language speech learning: The role of language experience in speech perception and production: A festschrift in honour of James E. Flege* (pp. 57-77). Amsterdam: John Benjamins.

Lee, J., Jang, J., & Plonsky, L. (2015). The effectiveness of second language pronunciation instruction: A meta-analysis. *Applied Linguistics* 36, 345–66.

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

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1993). Training listeners to perceive novel phonetic categories: How do we know what is learned? *Journal of the Acoustical Society of America, 94,* 1148-1151.

Munro, M. J., & Derwing, T. M. (2008). A developmental study of Mandarin and Russian speakers' English vowels. *Language Learning, 58*, 479-502.

Munro, M. J., Derwing, T. M., & Thomson, R. I. (2015). Implications of naturalistic segment development for pronunciation teaching technology. *International Review of Applied Linguistics in Language Teaching, 53*, 39-60.

Schmidt, R. (2001). Attention. In P. Robinson. (Ed.). *Cognition and second language instruction* (pp. 3-32). Cambridge: Cambridge University Press.

Thomson, R. I. (2013). Phonology: Quantitative methods. In C. A. Chapelle. (Ed). *The Encyclopedia of Applied Linguistics.* (pp. 4434-4444). Oxford: Wiley Blackwell. DOI: 10.1002/9781405198431.wbeal0912

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

Thomson, R. I. (2012b). *English Accent Coach* [Computer program]. Version 2.3

Thomson, R. I. (2011). Computer Assisted Pronunciation Training: Targeting second language vowel perception improves pronunciation. *CALICO Journal*, *28,* 744-765.

Thomson, R. I. & Derwing, T. M. (2015). The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics, 36*(3): 326-344.

Thomson, R. I., & Isaacs, T. (2009). Within-category variation in L2 English vowel learning. *Canadian Acoustics, 37,* (3), 138-139. Proceedings of the annual conference of the Canadian Acoustics Association.

Durham, K., Hayes-Harb, R., Barrios, S., Showalter, E. C. (2016). The influence of various visual input types on L2 learners' memory for the phonological forms of newly-learned words. 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. 98-107). Ames, IA: Iowa State University.

THE INFLUENCE OF VARIOUS VISUAL INPUT TYPES ON L2 LEARNERS' MEMORY FOR PHONOLOGICAL FORMS OF NEWLY-LEARNED WORDS

Kristie Durham, Department of Linguistics, University of Utah
Rachel Hayes-Harb, Department of Linguistics, University of Utah (contact author)
Shannon Barrios, Department of Linguistics, University of Utah
Catherine E. Showalter, Department of Linguistics, University of Utah

Showalter and Hayes-Harb (2013) reported that native English speakers who were exposed to tone marks (e.g., <gí>) outperformed learners not exposed to tone marks (e.g., <gi>) during a word learning task involving a Mandarin-like mini lexicon. The present study investigated whether the word form learning benefit reported in Showalter and Hayes-Harb (2013) is necessarily orthographic. Subjects were assigned to word learning conditions providing the pseudo-Mandarin stimuli from Showalter and Hayes-Harb in addition to various other types of visual information: screen color, screen position, and combinations of position, color, and letters. Results from Showalter and Hayes-Harb were not replicated. Text position was found to provide a benefit over other non-orthographic visual information (color), as well as orthographic information (tone marks). The findings suggest that orthography, while a likely contributor to a performance benefit, is not the only beneficial visual information during word learning. At this time, there is a need for studies that investigate the benefit of other types of visual information.

INTRODUCTION

Recent evidence suggests that written input can powerfully influence the acquisition of second language words and their phonological forms. It has been found that "misleading" written forms can cause learners to misremember the auditory forms of words (e.g., Young-Scholten, 2002; Brown, Hayes-Harb & Smith, in prep; Hayes-Harb, Nicol & Barker, 2010). For example, Hayes-Harb et al. (2010) taught native English speakers an artificial second language (L2) lexicon in two conditions: one where subjects were exposed to spelled forms that were congruent with English grapheme-phoneme correspondences (e.g., hear [fɑʃə]; see <fasha>), or one where the spelled forms were incongruent with English grapheme-phoneme correspondences (e.g., hear [fɑʃə]; see <faza>). At test, subjects in the incongruent condition were more likely to misremember the pronunciations of the newly-learned words—that is, these subjects misremembered the word [fɑʃə] as [fɑzə]. They interpreted this finding as indicating that subjects' memory for the newly-learned word's phonological forms was negatively impacted by the written input when the native language (L1) and the (artificial) L2 differed in grapheme-phoneme correspondences. In addition, Brown et al. (in prep) demonstrated that access to spelled forms in the L2 input can interfere with native English speakers' acquisition of German final obstruent devoicing. They taught native English speakers

"German-like" words in two conditions: one in which subjects saw spelled forms (e.g., hear [krɑt]; see <krad>), and one where no spelled forms were provided. At test, participants who saw <krad> misremembered the word as [krɑd], suggesting that in cases where auditory forms and written forms conflict, written input may override the auditory input.

On the other hand, we have also seen cases where learners exposed to words' written forms can exhibit more accurate memory for the words' phonological forms than learners who do not see written forms (e.g., Escudero, Hayes-Harb & Mitterer, 2008), though the literature also contains a number of counterexamples (Simon, Chambless & Alves, 2010; Showalter & Hayes-Harb, 2015; Hayes-Harb & Hacking, 2015).

Escudero et al. (2008) provided evidence that native Dutch speakers were more likely to have established contrastive lexical representations for newly-learned English-like words distinguished by /æ/ and /ɛ/ when they saw the words' spellings (the letters <a> and <e>) than when they did not. Showalter & Hayes-Harb (2013) explored whether such facilitative effects depend crucially on the familiarity of the written symbols to the learners—they asked whether unfamiliar but nonetheless systematic written symbols can be used by learners to accurately remember the phonological forms of words. In this study, 26 English speakers learned a set of eight Mandarin-like words differentiated by lexical tone. The words involved two sequences of segments ([fian] and [gi]) with four tones each (tones 1,2,3,4) to form eight distinct "words". Tone 1 is a high-level tone, tone 2 is high-rising, tone 3 is low-falling-rising, and tone 4 is high-falling. Each word was paired with a nonobject "meaning" (see Figure 1); the (randomly-assigned) connection of a phonological form and a meaning was intended to constitute a lexical item. Subjects were randomly assigned to one of two word learning conditions, distinguished by the presence/absence of unfamiliar diacritic lexical tone marks on Pinyin (Romanized) written forms (e.g., <gi> vs. <gí>). During a word learning phase, each auditory word and its corresponding image (accompanied by a written form) was presented to subjects who were simply instructed to "learn the new words and their meanings".
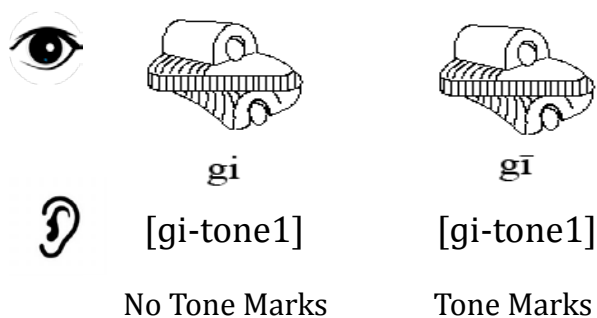

Figure 1. Example word learning trials, by word learning condition (Showalter & Hayes-Harb, 2013).

The word learning phase, which consisted of eight presentations of each word presented in a random order, was followed immediately by a criterion test. In the criterion test, a two-way forced-choice auditory word-picture matching task, subjects heard an auditory

form, saw a picture, and were asked to indicate by button press whether the word and picture (no longer accompanied by the written form) matched. The criterion test was intended only to ensure that subjects had reached a minimum level of word learning, and thus focused only on subjects' ability to distinguish [fian] words from [gi] words, but not minimal tone pairs (e.g., [gi-tone 1] from [gi-tone 2]). The criterion test involved 32 items—16 matched items (e.g., the picture that subjects were taught to associate with [gi-tone 1] presented along with the auditory form [gi-tone 1]). Subjects repeated the word learning – criterion test sequence until they reached 90% accuracy on the criterion test. The final test was identical to the criterion test except that subjects now were asked to distinguish between the minimal tone pairs. Showalter & Hayes-Harb (2013) found a significant effect of word learning condition on final test performance, with subjects in the Tone Marks condition exhibiting higher accuracy than subjects in the No Tone Marks condition. They concluded that the availability of orthographic tone marks in the input can help native English speakers learning pseudo-Mandarin remember lexical tone, and crucially, that even these unfamiliar written symbols influenced L2 word form learning.

Research Question

Given Showalter & Hayes-Harb's (2013) finding that native language familiarity with the specific written symbols is not a precondition for the symbols to influence novel word form learning, we now ask whether the word form learning benefit associated with the availability of tone marks reported by Showalter and Hayes-Harb (2013) is necessarily orthographic in nature. Specifically, we ask: To the extent that written forms can provide systematic visual cues to L2 phonological contrasts, might non-orthographic systematic visual cues similarly support novel word form learning? To answer this question, we conducted a replication and extension of Showalter & Hayes-Harb (2013), with a number of new word learning conditions in which each of the four lexical tone contrasts co-occur reliably with particular types of visual information, such as computer screen colors, screen position of images representing words' meanings, and screen position of words' written forms.

METHODS

All auditory and visual stimuli, procedures, and equipment used in the present study are identical to those used in Showalter and Hayes-Harb (2013). In addition to replicating the original two conditions (Tone Marks and No Tone Marks, which we now refer to as 'Image + Letters + Tone Marks' and 'Image + Letters', respectively), we created four new word learning conditions. These conditions involved manipulating two non-orthographic visual variables: screen position and screen color. The aim was to provide new types of systematic visual cues that could readily be presented on a computer screen. In the 'Image + Letters + Color' condition, subjects saw a picture and the letters only (i.e., without tone marks; e.g., <gi>), with the screen background color differing depending on the tone (tone 1: green, tone 2: yellow, tone 3: pink, tone 4: turquoise). In this way, the systematic visual information concerning tone that was provided to subjects in this condition was encoded in the screen color. In the 'Image + Letters + Position' condition, the only systematic visual cue to tone was encoded in the corner of the screen in which the letters (without tone marks) appeared. The picture was presented in the

center of the screen, while the positions of the written forms varied (tone 1: top left, tone 2: top right, tone 3: bottom left, tone 4: bottom right). In the 'Image + Position' condition, no letters were presented, and the systematic cue to tone was encoded in the position of the picture (tone 1: top left, tone 2: top right, tone 3: bottom left, tone 4: bottom right). Finally, we included the 'Image' condition, where only the image appeared, in the center of the screen, with no systematic visual cue to tone. The six word learning conditions (four new plus the two original Showalter & Hayes-Harb (2013) conditions) are summarized in Table 1. Example visual stimuli from each word learning condition are included in the Appendix.

Table 1
The Six Word Learning Conditions (S&HH = Showalter & Hayes-Harb, 2013)

Word Learning Condition	Orthographic		Non-Orthographic	
	Letters	**Tone Marks**	**Screen Position**	**Screen Color**
Image + Letters + Color *Systematic info: Screen color*	✔			✔
Image + Letters + Position *Systematic info: Position of text*	✔		✔	
Image + Position *Systematic info: Position of image*			✔	
Image + Letters + Tone Marks *Systematic info: Tone marks (S&HH)*	✔	✔		
Image + Letters *Systematic info: None (S&HH)*	✔			
Image *Systematic info: None*				

RESULTS

To determine the effect of word learning condition on subjects' ability to distinguish between the minimal tone pairs, we first converted the proportion correct scores to d' scores (see Figure 2). Next, we looked at the relationship between the number of word learning cycles required to reach criterion (see Table 2). The d' scores and number of word learning cycles were significantly correlated (Pearson r = .299, p= .001, n= 120); we thus included word learning cycles as a covariate in subsequent analyses. The d' scores were first submitted to an ANCOVA with word learning condition (six levels) and number of word learning cycles as a covariate. The main effect of word learning condition was not significant ($F_{(5,114)}=1.330$, p=.256, partial eta squared=.055). Because we were interested in the difference in performance between pairs of word learning conditions, we followed up with a number of planned comparisons. The significant pairwise comparisons are: Image + Letters + Position > Image + Letters +

Color (p=.013), Image + Letters + Position > Image + Letters (p = .030), and Image + Letters + Position > Image + Letters + Tone Marks (p = .012). These are indicated with asterisks in Figure 2. Thus, the only word learning condition that led to significantly more accurate performance than any others was the Image + Letters + Position condition.

Table 2
Mean Word Learning Cycles Required to Pass the Criterion Test, by Word Learning Condition (standard deviation)

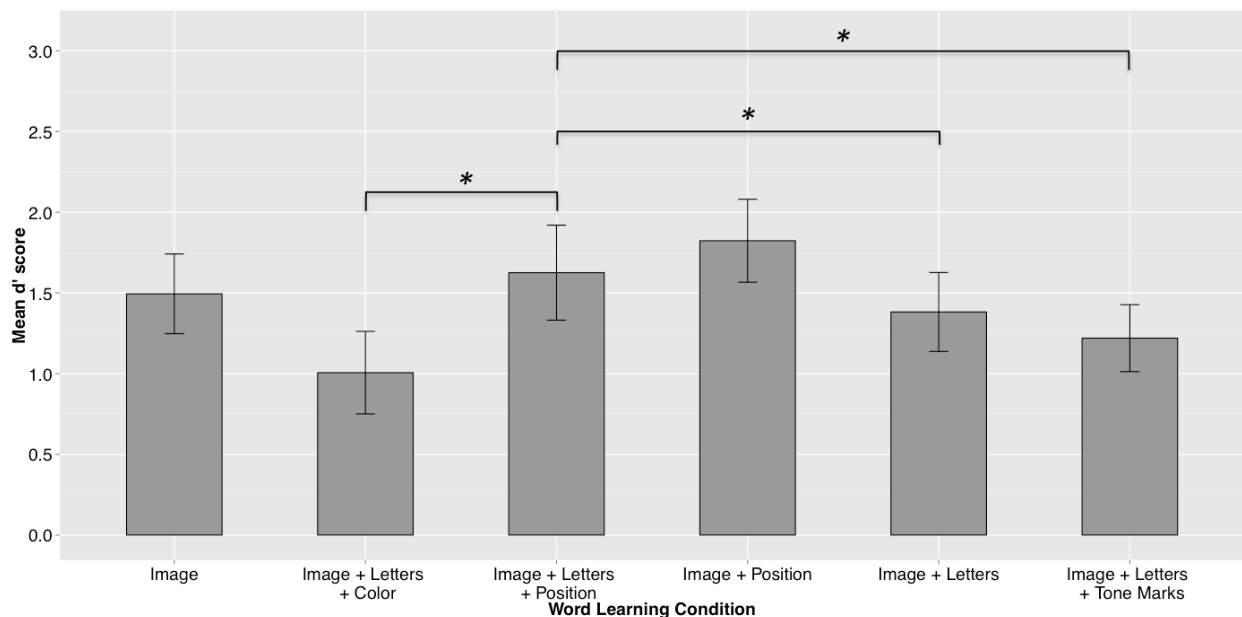| Word Learning Condition	Mean Word Learning Cycles	Mean d'
Image + Letters + Color	1.60 (1.00)	1.006 (1.14)
Image + Letters + Position	1.95 (.83)	1.626 (1.32)
Image + Position	2.50 (1.05)	1.82 (1.15)
Image + Letters + Tone Marks	1.85 (.75)	1.22 (0.93)
Image + Letters	1.80 (.83)	1.38 (1.09)
Image	2.50 (1.40)	1.50 (1.10)

Figure 2. Mean d' scores by word learning condition; whiskers represent one standard error of the mean; asterisks indicate significant pairwise comparisons.

The comparison between the Image + Letters + Tone Marks and the Image + Letters conditions (those from Showalter & Hayes-Harb, 2013) was not significant; we thus did not replicate the previous finding that the availability of tone marks in the written input facilitated the acquisition of lexical tone contrasts by native English speakers.

DISCUSSION

Recall that our research question was: Is the word form learning benefit associated with the availability of tone marks reported by Showalter and Hayes-Harb (2013) necessarily orthographic in nature? Of the six word learning conditions in the present study, the only one that resulted in significantly higher d' scores than other conditions was the 'Image + Letters + Position', with subjects' performance significantly higher than in the 'Image + Letters + Colors', 'Image + Letters' and the 'Image + Letters + Tone Marks' conditions. Why subjects in this condition performed so well, or indeed why these subjects did not also outperform subjects in the remaining two conditions, is unclear. However, that subjects in this condition outperformed, in particular, participants in the 'Image + Letters + Tone Marks' suggests that the answer to our research question is "no". The more accurate performance with the availability of a non-orthographic visual cue as compared to tone marks indicates that the benefit experienced by learners exposed to systematic visual information does not crucially rely on that information being orthographic in nature. In this case, we have provided evidence that non-orthographic systematic visual cues similarly support novel word form learning, at least in the case of lexical tone learning under the present study conditions.

We are left with the question of why only one of the non-orthographic visual cues conditions resulted in significantly more accurate performance than other conditions. First, why did the position of the **word** (Image + Letters + Position) but not position of the **picture** (Image + Position) result in a word form learning advantage? It is worth noting that in fact, descriptively, subjects in the Image + Position condition had the most accurate performance overall, though their performance was not significantly more accurate than that of any others in the ANCOVA, when the number of word learning cycles were taken into account. While the performance of this group was high, so was the mean number of word learning cycles required to pass the criterion test (mean = 2.5 cycles in this condition). It is also unclear why the position of the letters and not the color of the screen led to more accurate performance, though it is noteworthy that Godfroid, Ryu and Lin (2015) also provide evidence that a color-related cue did not positively impact native English speakers' acquisition of Chinese lexical tone. Future research may help to clarify theses questions, and may consider additional non-orthographic visual cues, as those explored here are not exhaustive of possible visual cues to novel phonological contrasts.

We also wish to stress that our current findings may challenge the robustness of written input effects, in that we did not replicate the Showalter and Hayes-Harb (2013) finding that subjects who had access to tone marks in the written input outperformed those who did not. In addition, while Showalter and Hayes-Harb did not find a correlation between the number of word learning cycles and accuracy at the final test, we found a significant

correlation in the present study. Further research, including replication studies, may clarify these issues.

In summary, we have provided evidence that the influence of written input on L2 word form learning may not be an exclusively orthographic effect. Specifically, we found that native English speaking participants exposed to non-orthographic by systematic visual (word position) cues to Mandarin lexical tone in fact outperformed participants who were exposed to orthographic (diacritic) cues. In this way, the present study contributes to an increased understanding of language learners' use of available cues to novel phonological contrasts in the input, and highlights the need for further research into the utility of various input types in supporting second language word learning.

ACKNOWLEDGMENTS

ABOUT THE AUTHORS

Kristie Durham is a graduate student in the Department of Linguistics at the University of Utah and member of the Speech Acquisition Lab. She received an M.A. in Languages and Literature at the University of Utah, where she developed an interest in second language pedagogy. Her current interests include second language phonological acquisition, as well as individual language learner differences.

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

Shannon Barrios (s.barrios@utah.edu) is an Assistant Professor and co-director of the Speech Acquisition Lab in the Department of Linguistics at the University of Utah. Her research uses a 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.

Catherine E. Showalter (catherine.showalter@utah.edu) is a doctoral student in the Department of Linguistics at the University of Utah. She currently works as a researcher in the Speech Acquisition Lab in the department. Her research has (thus far) focused primarily on the interaction of orthography and phonology in tasks completed by adult second language learners. Current research involves grapheme-phoneme correspondences and investigation of perceptual adaptability. Her other interests include second language phonological acquisition, non-native phonotactic acquisition, and creation of the lexicon.

REFERENCES

Brown, K., Hayes-Harb, R., & Smith, B. L. (In prep). The effects of orthographic input on the acquisition of German word-final devoicing by native English speakers.

Escudero, P., Hayes-Harb, R., & Mitterer, H. (2008). Novel second-language words and asymmetric lexical access. *Journal of Phonetics, 36*, 345-360.

Godfroid, A., Ryu, C., & Lin, C. (2015). Colorful benefits? The efficacy of dual coding in an online L2 Chinese tone perception study. Poster at the *Pronunciation in Second Language Learning and Teaching* conference, Dallas, TX, October 2015.

Hayes-Harb, R. & Hacking, J. F. (2015). The influence of written stress marks on native English speakers' acquisition of Russian lexical stress contrasts. *Slavic and East European Journal, 59*, 91-109.

Hayes-Harb, R., Nicol, J., & Barker, J. (2010). Learning the phonological forms of new words: Effects of orthographic and auditory input. *Language and Speech, 53*, 367–381.

Showalter, C. E., & Hayes-Harb, R. (2013). Unfamiliar orthographic information and second language word learning: A novel lexicon study. *Second Language Research*, *29*, 185-200.

Showalter, C. E., & Hayes-Harb, R. (2015). Native English speakers learning Arabic: The influence of novel orthographic information on second language phonological acquisition. *Applied Psycholinguistics, 36*, 23-42.

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.

Young-Scholten, M. (2002). Orthographic input in L2 phonological development. In P. Burmeister, T. Piske, & A. Rohde (Eds.), *An integrated view of language development: Papers in honour of Henning Wode* (pp. 263-279). Trier: Wissenschaftlicher Verlag Trier.

APPENDIX

The following are example visual stimuli from each of the six word learning conditions (for [fian-tone 1]).
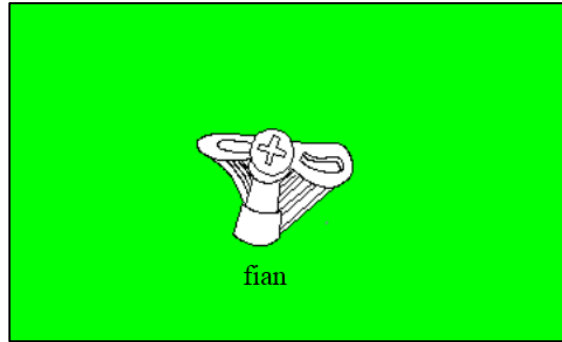
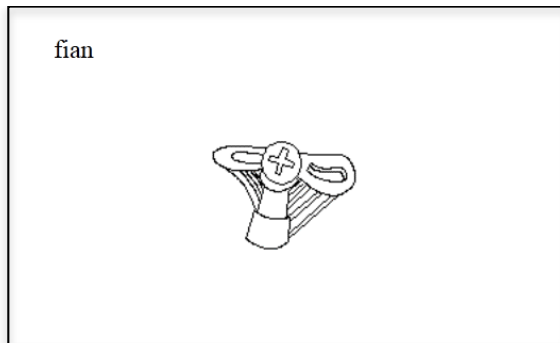Image + Letters + Color
(green background)
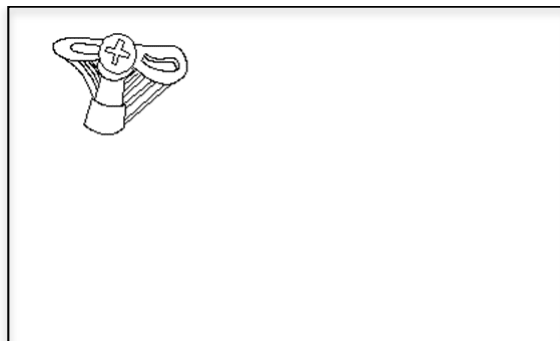

Image + Letters + Position
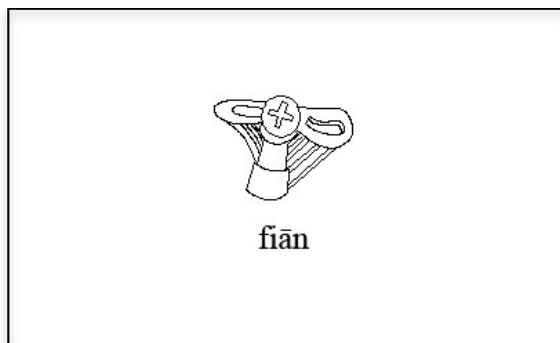

Image + Position

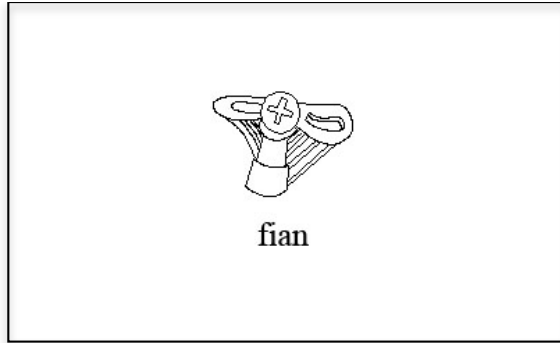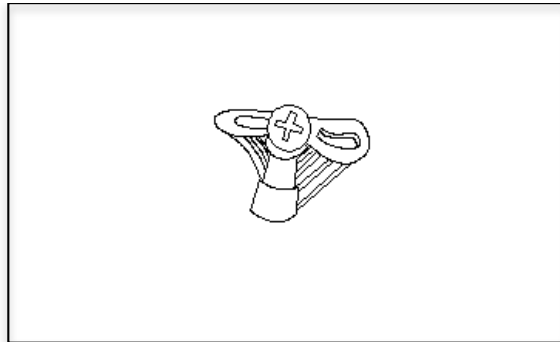

Image + Letters + Tone Marks

Image + Letters

fian

Image Only

BRAZILIAN ENGLISH x BRAZILIAN PORTUGUESE: A DYNAMIC APPROACH FOR THE ANALYSIS OF DIPHTHONGS IN FORENSIC CONTEXTS

Maria Lucia de Castro Gomes, Universidade Tecnológica Federal do Paraná

This research proposes different objectives and purposes from traditional analyses about special characteristics of the pronunciation of Brazilian speakers of English. Instead of the educational context, the arena is the forensic context, where a Brazilian would be the suspect of a crime and, supposedly, his/her voice in English would be the evidence. The aim of this article is to demonstrate preliminary results of an experiment using acoustic analysis, traditionally used for speaker comparison in forensic phonetics. The voices of Brazilians and Americans were recorded speaking English and Portuguese, and the diphthong [aɪ] was compared in three perspectives – inter-speaker, intra-speaker and inter-language. Preliminary results demonstrate that Brazilians do pronounce diphthongs in a way that makes them different from American native speakers, with values of F2 always higher in the beginning and in the end of the diphthong. As for duration, Brazilians present a greater relative duration of the transition (approximant) when compared to the vowel nucleus.

INTRODUCTION

Forensic phonetics is a branch of applied linguistics that provides knowledge and methods of analysis in the context of criminal investigations involving the use of spoken language. According to Nolan (1997), "since the recording of speech has become commonplace, law enforcers and courts have increasingly called on specialists to give opinions on speech samples, either in court or at the investigation stage" (p.746). The most common activities performed by professionals in this area are speaker comparison for identification, speaker profiling, determination of origin, design of voice line-ups, analysis of background noise and analysis of sound edition. With forensic phonetics as the research arena, this project explores a context in which a Brazilian would be the suspect of a crime and his/her recorded voice would be the evidence.

METHODS

This work presents preliminary results of an experiment on the characteristics of Brazilian speakers of English in the forensic context, using speaker comparison as a methodology. The segment for acoustic analysis was the diphthong. The participants were three Brazilians and three American female speakers, who are bilingual in Brazilian Portuguese and American English. Before the recording session, they received instructions on the research and signed to permit the use of their voice in the analyses.

The participants' voices were recorded reading carrier sentences[1] that contained the diphthong [aI] in target words, both in English and Portuguese. The words in English were *pie, buy, lie and high*, and the words (or nonce words) in Portuguese were *pai, bai, LAI and Rai*[2]. The carrier sentences were "Say _____ to me" and "*Diz _____ também*".

The three Brazilian participants were university students at the time of the study, two 19 year-olds and a twenty year-old. All three of them are from Curitiba, the capital of the southern state of Paraná, and had always lived there. They have Portuguese as their native language and are fluent in English (they said they spoke English every day, mainly at the university). One of them is also fluent in German. The three American participants – two 23 year-olds and a 25 year-old – were English Teaching Assistants sponsored by the Fulbright Program at a university in Curitiba. They were born in the USA, New York (NY), Los Angeles (CA), Lyndhurst (OH), and the three had lived in other cities of the United States. One of them, of Japanese ancestry, had also lived in Tokyo and Honolulu. Another participant of eastern ancestry, Vietnamese, had never lived outside the US. They all had English as their native language, were fluent in Portuguese, and in other languages as well. One was fluent in Japanese and Spanish, the other in Vietnamese, and the third, in French.

The hypothesis that motivated this research is that Brazilian speakers of English may have special features that might indicate their origin in speaker comparison cases in forensic contexts. The choice for the diphthong as the object of analysis came from the potential of vowels in forensic analysis, as advocated by several authors (Rose, 2006; Mcdougall, 2002, 2004). The measurement of F2 was inspired by Nolan and Grigoras (2005), who claimed that the dynamics of the trajectory of formants in diphthongs is very useful in speaker recognition. According to the authors, the second formant presents a clearer trajectory. Haupt and Seara (2012) found a clear transition region between the two vowel targets of the diphthong [aɪ], observed in the F2 trajectory. Grigoras and Nolan (2005) also state that F2 frequency analyses may be more interesting than the F1 frequency and the F1 values may present errors in telephone recordings due to bandwidth problems.

Preliminary results
The results of the F2 measure come from five repetitions of four words in carrier sentences, in Portuguese and English, pronounced by each of the six participants. For duration, only five repetitions of the words *pai* and *pie*, by the six participants, will be described. Table 1 demonstrates F2 measures of the start and end points of the diphthongs.

---

[1] There were two other tasks (text reading and a short conversation), but their results won't be shown in this article.
[2] From those four words, only *pai,* which means *father*, is a real word in Portuguese. The others were invented and inserted in texts. The texts were designed to create a context, so we could have the same pronunciation in English and Portuguese. Brazilians usually use the term *bai, bai* to say goodbye, copying the English expression *bye-bye*, so *bai* was inserted in a text, under this meaning. LAI is an acronym, meaning *Lei de Acesso à Informação* – Law on the Access to Information, and the text had the acronym with that meaning. Finally, *Rai* was in the text as a short name for a girl named *Raiana*.

Table 1

F2 measures at the start and end points of the diphthong

Word	Participant	F2-START	F2-END	Word	Participant	F2-START	F2-END
	BRAZ	1490.667	2383.867		BRAZ	1560.867	2521.267
PAI	AM	1432.333	2275.467	PIE	AM	1292.333	2362.267
	BRAZ	1541.6	2358.933		BRAZ	1557.4	2585.467
BAI	AM	1431.733	2314.933	BUY	AM	1340.6	2346.733
	BRAZ	1411.467	2305.533		BRAZ	1364.2	2556.33
LAI	AM	1359.4	2286.667	LIE	AM	1272.67	2302.27
	BRAZ	1651.067	2297.2		BRAZ	1641.067	2565.133
RAI	AM	1559.4	2326.933	HIGH	AM	1399.867	2365.8

Statistical analyses of the data are yet to be done, but it is already possible to see some tendencies in the results. As seen in Table 1, the means of F2 values, either at the beginning or at the end of the diphthong (except for *Rai* F2-Start in Portuguese), had higher values in the productions by the Brazilian subjects, both in Portuguese and English words. Figure 1 provides a better view of the results.
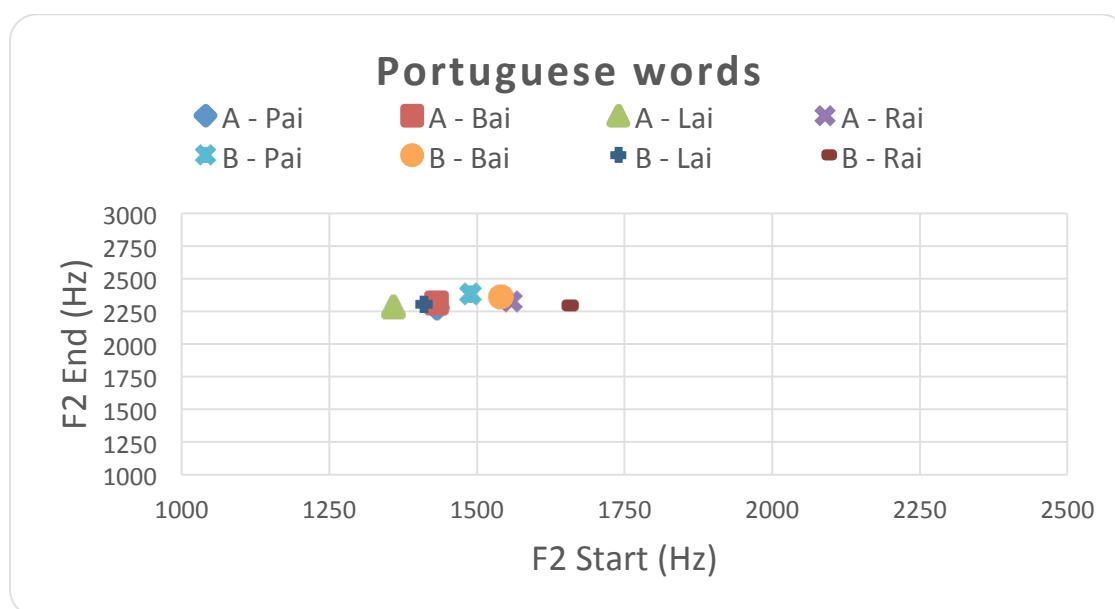

Figure 1. F2 converging point at the beginning and end of the Portuguese diphthongs [aɪ] produced by the Americans (A) and the Brazilians (B) – Portuguese words.

For the words in Portuguese, we can see that the mean F2 at the beginning of the diphthong is between 1,359 Hz (in *LAI* by the Americans) and 1,651 Hz (in *Rai* by the Brazilians). On the other hand, F2 at the end of the diphthong is between 2,275 Hz (in *pai* by the Americans) and 2,383 (in *pai*, by the Brazilian). We can see a larger difference in F2 values from the start point to the end point of the diphthong. We can also see that there may be a phonological influence, not analyzed here. However, we can see that at the start point, F2 values are higher when the preceding consonant is a plosive. And we can also see that F2 values in the beginning are higher when the diphthong is preceded by [h] than when preceded by [l], when pronounced by both the Brazilians and the Americans. Comparing the start point of F2 produced by both groups, word by word, one can clearly see the converging point always further to the right when produced by the Brazilians. The difference between Brazilian and American productions is also present in the English words.

Figure 2. F2 converging point at the beginning and end of the Portuguese diphthongs [aɪ] produced by the Americans (A) and the Brazilians (B) – English words.

In Figure 2 we can also observe that the converging points of F2 produced by the Brazilians have higher F2 values. The mean F2 value at the start point of the diphthong is between 1,272 Hz (in *lie* by the Americans) and 1,641 Hz (in *high* by the Brazilians). The mean F2 value at the end of the diphthong is between 2,302 Hz (in *lie* by the Americans) and 2,585 (in *buy* by the Brazilians). Comparing word-by-word, it is possible to observe that the converging points of the two F2 measures are always at a higher position at the Brazilians' graph data, the values of the end point of F2 being about 200 Hz higher. According to Yavas (2011, cited by Woolums, 2012), in order to qualify as significant, the difference in F2 must be 200 Hz.

As for the Brazilians' productions of Portuguese and English words, there is a tendency for higher value of F2 at the end of the diphthongs in the words in English, as can be seen in Figure 3. Except for the word *pie,* which had a difference of 137.4 Hz higher in the F2 value at the end of the diphthong, all the other differences were higher than 200 Hz.
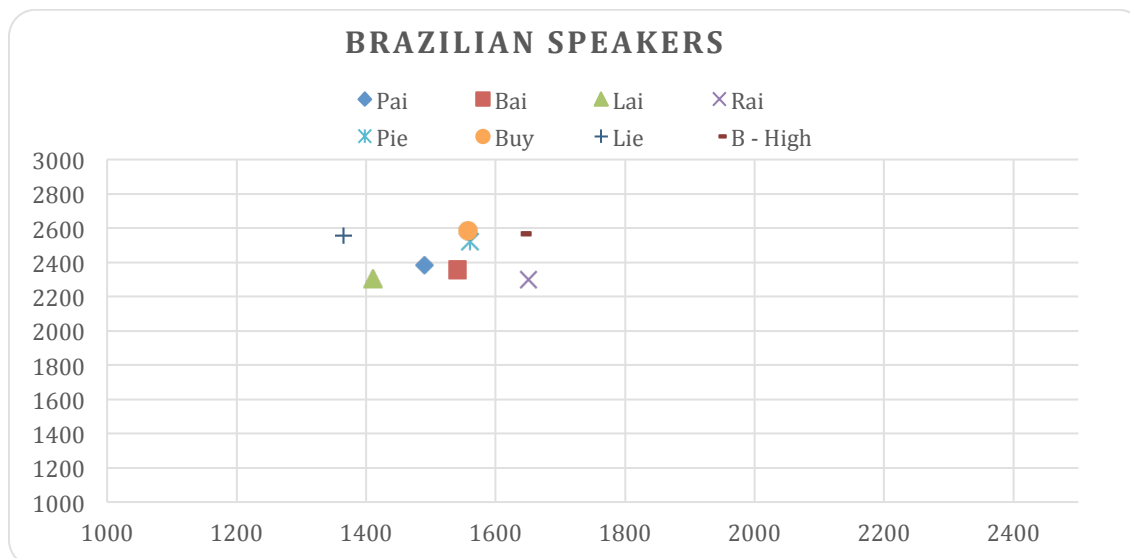
Figure 3. F2 converging point at the beginning and end of the diphthong [aɪ] produced by the Brazilians in the words in Portuguese and in English.

Figure 3 demonstrates that the converging point of F2 at the beginning and at the end of the diphthong is always higher for the words in English than the words in Portuguese. After comparing the diphthongs produced by the Brazilian and American participants in the words in English and Portuguese, and between the diphthongs in the words in Portuguese and in English produced by the Brazilians, the pair of words *pai/pie* was chosen to establish a comparison between the three informants of each nationality. Checking some dispersion may indicate important linguistic variations, mainly due to dialect differences. The first charts compare the words pronounced in the native language, i.e. Portuguese by the Brazilians and English by the Americans.
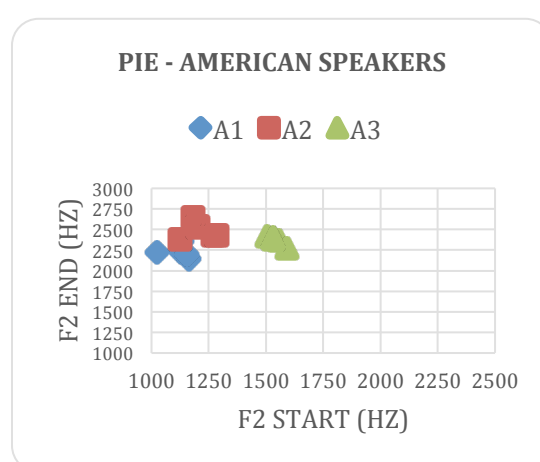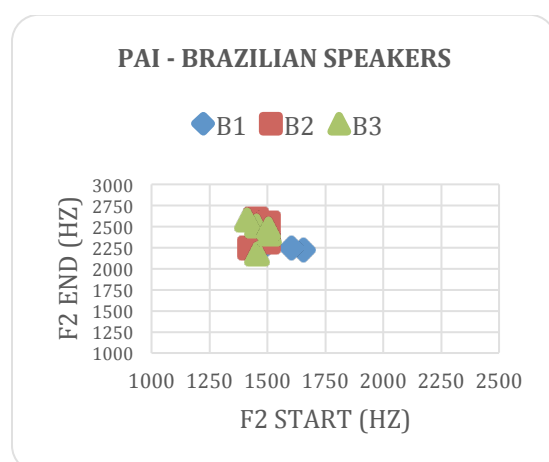
Figure 4 and 5. F2 converging point at the beginning and end of the diphthongs [aɪ] by the Brazilians (Figure 4) in the word *pai* and by the Americans (Figure 5) in the word *pie*.

The charts indicate that the productions of the Brazilians exhibit less variation than those of the Americans. There may be less influence of dialectal variation since the Brazilian subjects were all born in the same city, while the Americans were all born in different regions of the US. Another fact that may be relevant is that the diphthong [aɪ] is undergoing raising in some Midwestern portions of the US (Woolums, 2012).

Analyses of the productions of the same pair of words as a foreign language, by both groups of subjects, is displayed in Figures 6 and 7.
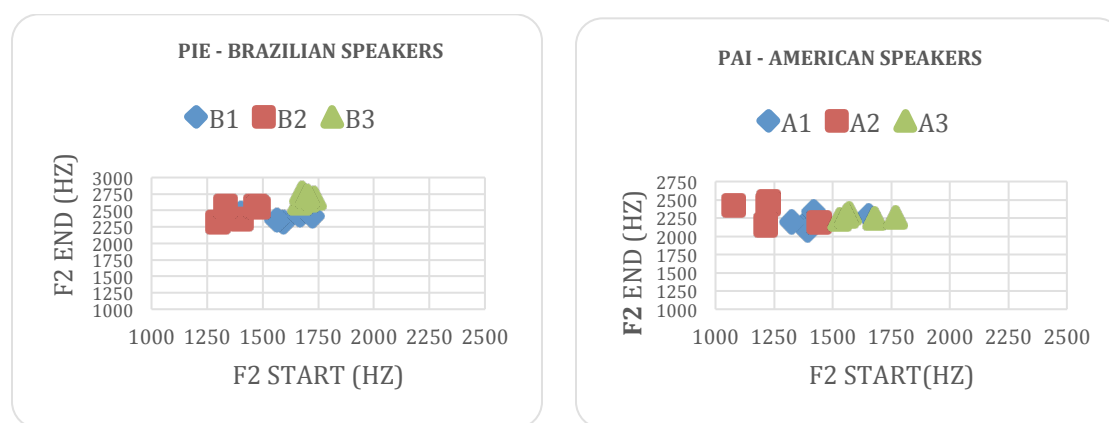
Figure 6 and 7. F2 converged points at the beginning and end of the diphthongs [aɪ] by Brazilians respondents (Figure 6) in the word *pai,* and Americans (Figure 7) in the word *pie.*

Duration

Next, we analyze the duration of the pair of words *pie/pai* in detail. For the duration of the vowel, the start and end points were the first and last periods that had considerable amplitude. From the point where the trajectory of formants (F1 and F2) changed, we considered the final point of the vowel and the start of the approximant. The end point of the approximant was the last pulse of the waveform. Figure 8 demonstrates an example of duration measurement using PRAAT. The first tier shows the waveform, the second tier shows the spectrogram, in the third tier we can see the word, and finally in the fourth, we see the labels of the segments. Although the consonant is not a focus of analysis here, it was also labeled, as it may be important in the word duration, especially if we consider the aspiration in the English word.
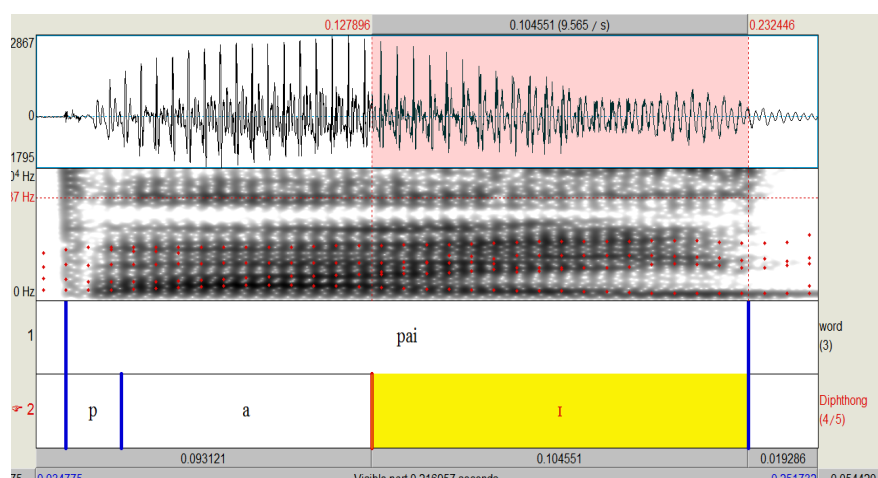
Figure 8. Word *pai* produced by Subject B3.

For the calculation of the relative duration of the vowel and the approximant, we divided the absolute duration of each by the absolute duration of the word, and the result was multiplied by 100, following Silva (2014). We can see, in the tables below, the mean values of the absolute duration of the word (in seconds) and the relative duration of the vowel and the approximant of the words *pai* and *pie* produced by the Brazilians and the Americans.

Table 2

Absolute duration of the word and relative duration of the vowel and the approximant by the Brazilians and the American, and standard deviation in the Portuguese word pai.

INF	Word dur.	Vowel rel.dur	Approx rel. dur.	INF	Word dur.	Vowel rel.dur	Approx rel. dur.
B1				**A1**			
	0.18	29.42	57.13		0.28	42.34	45.95
B2				**A2**			
	0.19	32.04	59.18		0.24	40.30	45.61
B3				**A3**			
	0.24	37.33	58.86		0.25	40.37	37.24
MEAN				**MEAN**			
	0.20	**32.93**	**58.39**		**0.26**	**41.00**	**42.93**
SD				**SD**			
	0.02	3.29	0..90		0.015	0.94	4.02

Table 3

Absolute duration of the word and relative duration of the vowel and the approximant by the Brazilians and the American, and standard deviation in the English word pie.

INF	Word dur.	Vowel rel.dur	Approx rel. dur.	INF	Word dur.	Vowel rel.dur	Approx rel. dur.
B1	0,26	35,42	43,17	A1	0,30	31,84	36,24
B2	0,24	42,49	51,20	A2	0,40	34,26	39,58
B3	0,29	34,58	51,31	A3	0,34	36,78	39,18
MEAN	**0,26**	**37,50**	**48,55**	MEAN	**0,35**	**34,30**	**38,33**
SD	0,011	3,53	4,01	SD	0,041	2,016	1,489

Table 2 shows the mean duration of the word *pai* [paɪ] in Portuguese, and Table 3 shows the values of the word *pie* [paɪ] in English. On the left side of the two tables are the values of the productions of the Brazilian subjects, and on the right, the ones of the American subjects. Comparing the absolute duration of the word, there is a higher mean duration of the word in English than in Portuguese by both groups of subjects. On the other hand, the mean duration of the word is higher when produced by the Americans, in both languages, Portuguese and English. The difference between the relative duration of the vowel and the approximant is higher when produced by the Brazilians, mainly in Portuguese.

DISCUSSION

Though very preliminary, the above data indicate that the hypothesis that Brazilians have special characteristics when they speak English is likely to be confirmed. F2 measures of the diphthong [aɪ] in the Brazilian subjects' productions reveal a different characteristic, both in comparison to the production of the Americans in English, and when compared to their own productions of words in English and in Portuguese.

It is important to acknowledge several factors that could certainly be at stake in the data analysis, mainly the consonantal context and dialectal variation. Concerning consonantal context, the influence of the consonant that precedes the diphthong is evident in the F2 values, especially at the beginning of the diphthong. Figure 3 clearly shows that influence, but it also points to a tendency for vowel fronting by the Brazilian subjects, mainly in the English words, since higher F2 values indicate greater fronting.

A second factor that might influence the results and should be taken into account is the dialectal variation of diphthongs in English. Jacewicz and Fox (2013) analyzed [aɪ] in three different dialects of North American English and claim that the duration of diphthongs varies significantly in different regions of the United States and these temporal factors may have profound effects on the trajectory of the formants along the vowel. Jacewicz and Fox (2013) claim that there is a

complex interaction between the duration of the vowel, the consonant context and the speaking style, influencing the frequency of formants.

Jacewicz and Fox (2013) was dedicated to temporal issues of vowels, and reveals that the duration differences due to consonant context interact with dialectal variation. One of the regional dialects analyzed was the one produced in Ohio, birth region of one of the participants in this study (A3). She had very different results from the other two American subjects. Her F2 values, especially on the starting point of the diphthong, were closer to the F2 values of the Brazilian subjects. The difference between the F2 values of the Brazilians and the Americans could have been larger if that informant was not in the group. Figure 4 shows that the converging points of A3, in the productions of the word *pie*, is closer to the Brazilians' data. As mentioned above, work by Woolums (2012) also demonstrates [aɪ] raising in some US dialects.

Analysis of the duration in these preliminary results may also indicate a difference in the pattern of production between the Brazilians and Americans. Tasko and Greilick (2010) analyzed the duration of American diphthongs, comparing clear and conversational speech, and concluded that although the duration of diphthongs is greater when spoken clearly, the ratio is approximately the same as in conversation, i.e. 46% in articulated speech and 45% in the conversation. Analyzing the production of diphthongs at different speeds, they tried to explain the phenomena outlined in the gestural overlap model designed by Tjaden and Weismer (1998 *apud* Tasko & Greilick, 2010), saying that "shorter overall durations can be

achieved by increasing the degree of gestural overlap, whereas increasing duration would be achieved by reducing the degree of gestural overlap" (Tasko & Greilick, 2010, p.97).

In this research, the Brazilian subjects' data present a greater relative duration of the approximant when compared to the vowel, demonstrating a greater overlap of the approximant. The Americans distributed the vocalic segments in a more balanced way.

ABOUT THE AUTHOR

Maria Lúcia de Castro Gomes is a PhD in Linguistics and works as a professor at UTFPR – Universidade Tecnológica Federal do Paraná, in Brazil. She coordinates the course of Language and Literature at that institution. Her research focus in on phonetics and phonology of English and Portuguese, and forensic phonetics.

REFERENCES

Becker, M. R. (2012). O construto "Inteligibilidade" da língua Inglesa sob o paradigma de língua franca. In *Anais do X Encontro do CELSUL – Círculo de Estudos Linguísticos do Sul* (pp. 1-10). Cascavel: Unioeste.

Becker, M. R. (2014). Globalização, Inglês como Língua Franca e Inteligibilidade. In: In Brawerman-Albini, A & Gomes, M. L. C. (Eds) *O jeitinho brasileiro de falar inglês –*

pesquisas sobre a pronúncia do inglês por falantes brasileiros (pp. 223-238). Campinas: Pontes Editora.

Brawerman-Albini, A., & Gomes, M. L. C. (2014a) Aquisição de inglês como segunda língua perspectivas teóricas para o ensino de pronúncia. In A. Brawerman-Albini & M. L. C. Gomes. (Eds). *O jeitinho brasileiro de falar inglês – pesquisas sobre a pronúncia do inglês por falantes brasileiros* (pp. 19-32). Campinas: Pontes Editora.

Brawerman-Albini, A., & Gomes, M. L. C. (2014b) *O jeitinho brasileiro de falar inglês pesquisas sobre a pronúncia do inglês por falantes brasileiros.* Campinas: Pontes Editora.

Cruz, N. F. C., & Pereira, M. A. (2006). Pronúncia de aprendizes brasileiros de inglês e inteligibilidade: um estudo com dois grupos de ouvintes. *Revista Virtual de Estudos da Linguagem – ReVEL*, *4*(7) 1–26.

Cruz, N. F. C. (2014). Inteligibilidade e o ensino de pronúncia do inglês para brasileiros. In A. Brawerman-Albini & M. L. C. Gomes. (Eds). *O jeitinho brasileiro de falar inglês – pesquisas sobre a pronúncia do inglês por falantes brasileiros* (pp. 239-252). Campinas: Pontes Editora.

Crystal, D. *English as a global language*, 2nd ed. (2010). Cambridge: Cambridge University Press.

Gold, E., & French, P. (2014). Commonly used forensic speaker comparison parameters: expectations, misconceptions and empirical findings**.** *International Association for Forensic Phonetics and Acoustics 2013 Annual Conference*. Retrieved from http://www.iafpa.net/conf.htm.

Gomes, M. L. C., Brawerman-Albini, A., & Engelbert, A. P. (2013). *The perception of vowel epenthesis and stress in an English as a Lingua Franca context. Proceedings of the International Symposium on the Acquisition of Language Speech*, Concordia Working Papers in Applied Linguistics.

Graddol, D. (2006). *English Next:* Why global English may mean the end of 'English as a Foreign Language'. Retrieved from www.britishcouncil.org/learning-research

Haupt, C., & Seara, I. C. (2012). Caracterização acústica do fenômeno de monotongação dos ditongos [ai, ei e oi] no falar florianopolitano. *Linguagem & Ensino*, *15*(1), 263-290.

Jacewicz, E., & Fox, R. A. (2013). Cross-dialectal differences in dynamic formant patterns in American English vowels. In *Vowel Inherent Spectral Change*, ed. by Geoffrey S. Morrison and Peter Assmann, 177-198. New York: Springer-Verlag.

Mcdougal, K. (2002). Speaker-specific formant dynamics: An experiment on Australian English /aɪ/. *Speech, Language and the Law, 11*(1).

Mcdougal, K. (2006). Dynamic features of speech and the characterization of speakers: towards a new approach using formant frequencies. *Speech, Language and the Law* 13(1).

McCrocklin, S. (2012). *The role of word stress in English as a lingua franca. Proceedings of the 3rd Annual Pronunciation in Second Language Learning and Teaching* (pp. 249-254). TESL/Applied Linguistics Iowa State University.

Munro, J. M., Derwing, T. M., & Holtby, A. K. (2012). *Evaluating individual variability in foreign accent comprehension. Proceedings of the 3rd Annual Pronunciation in Second Language Learning and Teaching Conference* (pp. 233-239). TESL/Applied Linguistics Iowa State University.

Nolan, F. (1995). Can the definition of each speaker be expected to come from the laboratory in the next decades? *ICPhS 95* Stockholm, 3, 130-137.

Nolan, F. (1997). Speaker recognition and forensic phonetics. *The Handbook of Phonetic Sciences* (pp. 744-767). Oxford: Blackwell.

Nolan, F. & Grigoras, C. (2005). A case for formant analysis in forensic speaker identification. *Speech, Language and the Law. 12*(2), 143-173.

Rauber, A.S., Watkins, M., Silveira, R., & Koerich, R., Eds. (2010). *The acquisition of second language speech: Studies in honor of Professor Barbara O. Baptista*. Florianópolis: Insular.

Roberson, A. (2012). *Analyzing comprehensibility among non-native speakers of English: The effect of listener first language background. Proceedings of the 3rd Annual Pronunciation in Second Language Learning and Teaching* (pp. 240-248). TESL/Applied Linguistics Iowa State University.

Rose, P. (2006). *The intrinsic forensic discriminatory power of diphthongs*. Proceedings of the 11th Australian International Conference on Speech Science & Technology, University of Auckland, New Zealand. December 6-8.

Schadech, T. S. & Silveira, R. (2014). A inteligibilidade e o ensino de pronúncia: interface entre pesquisa e pedagogia. In A. Brawerman-Albini & M. L. C. Gomes. (Eds). *O jeitinho brasileiro de falar inglês – pesquisas sobre a pronúncia do inglês por falantes brasileiros*. Campinas: Pontes Editora. 253-271.

Silva A. H. P. Organização temporal de encontros vocálicos no Português Brasileiro e a relação entre Fonética e Fonologia. *Letras de Hoje*, Porto Alegre, v. 49, n. 1, p. 11-18.

Tasko S. M., & Greilick K. (2010). Acoustic and articulatory features of diphthong production: A speech clarity study. *Journal of Speech, Language, and Hearing Research*, *53*, 84–99.

Woolums, N. (2012) Phonetic manifestations of /aɪ/ raising. *Linguistic Portfolios*, *1*, Art.19.

Johnson. O. D., Kang, O., & Ghanem, R. (2016). 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*, ISSN 2380-9566, Dallas, TX, October 2015 (pp. 119-129). Ames, IA: Iowa State University.

LANGUAGE PROFICIENCY RATINGS: HUMAN VS. MACHINE

David O. Johnson, Northern Arizona University
Okim Kang, Northern Arizona University
Romy Ghanem, Northern Arizona University

This paper explains a computer model that mechanically assesses the verbal proficiency of audio recordings of unconstrained non-native English speech. The computer model utilizes machine learning and eleven suprasegmental measures split into four categories (stress, pitch, pause, and temporal) to compute the proficiency levels. In an experiment with 120 non-native English speaker's monologs from the speaking section of the Cambridge ESOL General English Examinations, the Pearson's correlation comparing the certified Cambridge English Language Assessment proficiency scores and the computer's computed proficiency scores was 0.718. This human-computer correlation is greater than that of other related computer programs (0.55-0.66) and is nearing that of human examiners (0.70-0.77) with regards to inter-rater reliability.

INTRODUCTION

Language proficiency assessments are intended to measure reading, writing, listening, and speaking abilities. Humans can score proficiency assessments; but they are costly to employ, train, and compensate; they take a long time to score assessments which produces postponements in providing the results to the candidates; and still with multiple raters, rubrics, and frequent inter-rater reliability testing, humans lack consistency and objectivity. For example, Kang and Rubin (2009) found that listener's attitudinal and background factors accounted for 18-23 % of the variance in human assessment. Innovations in artificial intelligence and natural language processing have resulted in computer programs that can automatically rate language proficiency. Automated scoring systems generate assessments quicker and more economically than human scoring and they are more consistent and equitable in scoring than humans. This is especially true with automated delivery and rating of reading, writing, and listening skills (Attali & Burstein, 2006; Burstein et al., 1998; Chodorow & Burstein, 2004; Landauer & Dumais, 1997; Rudner, Garci, & Welch, 2006; Zechner, Higgins, Xi, & Williamson, 2009). Automated reading and listening assessments are characteristically multiple-choice. They are simple to create and manage, comparatively uncomplicated to grade mechanically, and substantiated by a robust foundation of assessment philosophy and statistical practices. Automated writing tests are usually delivered online and scored automatically. They are written constructed response items where the examinees write a succession of compositions on designated subjects.

Automatic Speaking Proficiency Assessment

Speaking skill assessment is more difficult than other assessments. There are two categories for automated scoring systems in speech: constrained and unconstrained (spontaneous). Constrained speech assessment is the easier of the two to automate. Typically test-takers are requested to respond orally to constructed response items like reading aloud, repeating sentences, building sentences, giving short answers to questions, or retelling brief stories. For some tasks, one correct word sequence is expected for each response. In other tasks, items can have multiple correct answers. The computer recognizes the words spoken with an automatic speech recognizer (ASR) and compares them to the hypothesized response (content). It locates linguistic units (segments, syllables, and words) and measures the pace, fluency, and pronunciation of those words in phrases and sentences (prosody). Then, the computer combines the content and prosodic measures using statistical modeling techniques and calculates an overall score as a weighted combination of the sub-scores. Their use in evaluating constrained speech proficiency has been confirmed by establishing that the automated scores were substantially correlated with those that human raters ascertained from speaking proficiency examinations (Bernstein, Van Moere, & Cheng, 2010).

Existing Automatic Unconstrained Speaking Proficiency Assessment

Unlike constrained speech, unconstrained speech is irregular and variable making automatic proficiency scoring of it more challenging. Asking candidates to converse on a subject for one or two minutes (e.g., what is happening in a picture) is the normal means of obtaining unconstrained speech samples to assess. SpeechRater[SM] is an instance of an operational computerized unconstrained English speech proficiency assessment tool (Zechner et al., 2009). As illustrated in Figure 1, SpeechRater[SM] detects the words in the candidate's speech with an ASR.
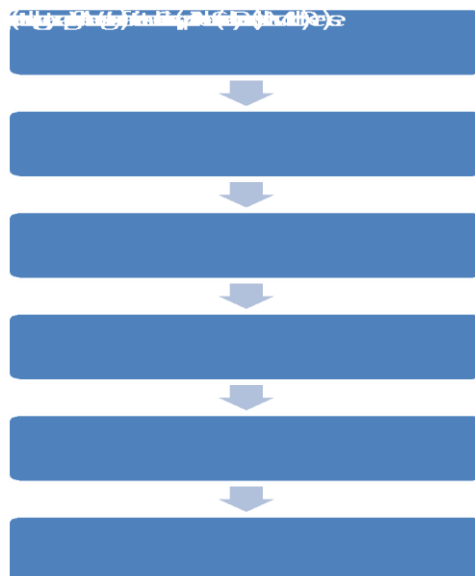
Figure 1. SpeechRater[SM].

It then uses the output from the ASR to compute eleven prosodic measures: average chunk length (in words), where a chunk is segment of contiguous words, articulation rate, mean deviation of chunks (in words), total duration of silent pauses divided by number of words, average silent pause duration (in seconds), average of long silent pause (greater than or equal to 500 ms) duration, frequency of long silent pauses divided by number of words, types of unique words per second, number of types divided by duration of entire transcribed segment exclusive of inter-utterance pauses, normalized global HMM acoustic model score, and normalized global language model score. The eleven measures are then combined with multiple-regression to estimate a speaking proficiency rank of one (lowest) to four (highest). The Pearson's correlation between the ranks assessed by a human and those estimated by SpeechRater^{SM} was 0.55. A Classification and Regression Tree (CART) machine learning version, which was not deployed, had a stronger correlation of 0.62.

New Computer Model for Automatically Scoring Unconstrained Speech Proficiency

We developed a computer model that automatically scores unconstrained English speech proficiency from suprasegmental measures derived from Brazil's (1997) prosody model. The computer programs calculate the suprasegmental measures from the output of an ASR that recognizes phones instead of words. As depicted in Figure 2, in contrast to the method that SpeechRater^{SM} employed, our method has three benefits.

Figure 2. Comparison of SpeechRater^{SM} and our method of automatic proficiency scoring of unconstrained speech.

The first benefit is a consequence of the ASR recognizing phones instead of words. This is because the ASR only has to recognize the relatively tiny number of phones that are used in English words as opposed to recognizing the hundreds of thousands of words that could appear in unconstrained speech. Since there is a lesser quantity of phones to recognize than words by several orders of magnitude, the phone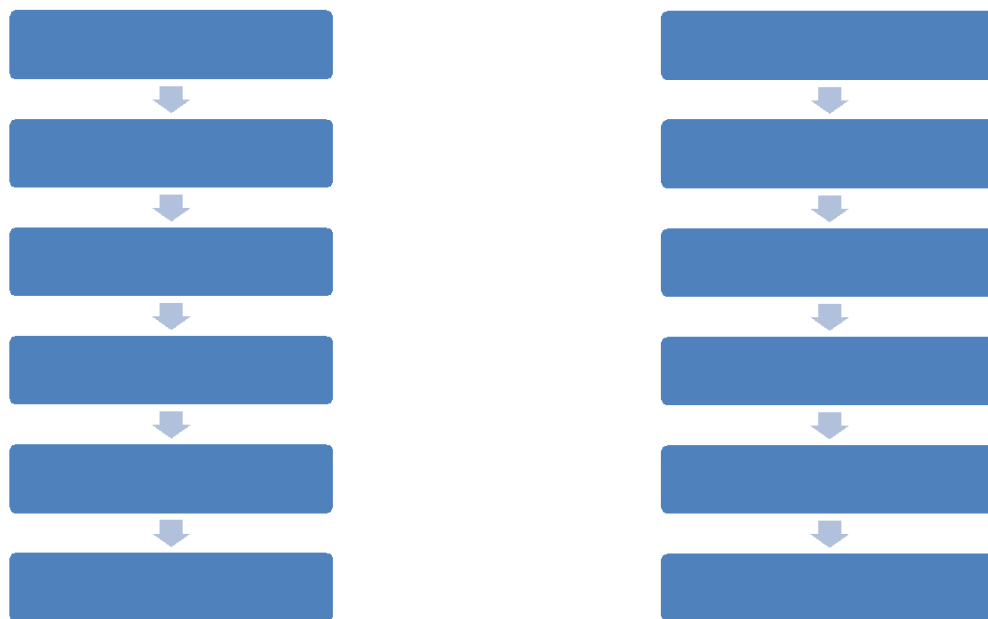 error rate (PER) of an ASR is predictably less than the word error rate (WER). This lower PER can lead to more correct proficiency scores. The second benefit of our tactic is making use of, along with fluency features, intonational measures drawn from a larger set of suprasegmental measures which were found to explain more than half of the variance in speaking proficiency scores (Kang, Rubin, & Pickering, 2010). Utilizing machine learning, which is a sub-domain of artificial intelligence, results in the third benefit of our approach. Machine learning normally achieves better results than multiple-regression. The fact that Zechner et al. (2009) built a machine learning version (i.e., CART) of SpeechRater[SM] which had a higher correlation between machine and human proficiency scores than their multiple-regression version is evidence of this.

We begin this paper with an overview of Brazil's (1997) prosody model and a description of the corpus and experimental methods we used to test the computer model we developed to automatically score the English proficiency of unconstrained speech. Then, we report the results and discuss them. We finish with a conclusion and some areas for further study.

METHODS

Brazil's Prosody Model

One of the earliest to put forward the notion of discourse intonation was Brazil (1997). He defined intonation as the linguistically deliberate variation of oral pitch intensity and duration throughout a discourse to relay information beyond that conveyed by the words and grammar. He held forth that the communication purpose of a discourse was realized by the recurring and purposeful selection of one pattern of intonation from an array of patterns. Brazil's model did not require additional phonological or acoustic classifications of the pitch attribute of speech which earlier intonation models had required. Nevertheless, his model assigned fresh inferences and connotations to orthodox intonation components (Chun, 2002). His model is regularly made use of in learning and teaching a language for the reason that it is founded on the use of intonation in a discourse to accomplish linguistic objectives that reach beyond the sentence level. He maintained that the four principal features of his model, i.e., tone unit, prominence, tone choice, and relative pitch, offered a practical structure for examining and studying the use of intonation that speakers exercised in a discourse. The main features of his model remain true for every facet of discourse; whether it is a dialog or a monolog consisting of either unconstrained or constrained speech.

Brazil characterized a tone unit as a fragment of a speech that a listener can perceive has an arrangement of falling and rising tones which is not the same as the arrangement of another fragment of the speech (Brazil, 1997). Then he stated that all tone units include a minimum of one prominent syllable. Chun (2002) added that syllables become prominent by being accentuated with extra pitch (fundamental frequency in Hz), intensity (amplitude in dB), duration (length in seconds), or a mixture of the three. Brazil insisted that prominence was ascribed to the

122

syllable, and not the word. Brazil differentiated prominence from lexical stress. Lexical stress is the normal, or dictionary defined, stress applied to syllables within a word. In opposition, prominence is the application of supplementary pitch, intensity, or duration on a syllable, even if it is lexically stressed, to call attention to a word's importance or to recognize its difference. The initial prominent syllable is called the key and the last is called the termination. A solitary prominent syllable in a tone unit is considered both the key and termination. The arrangement of falling and rising intonation of a tone unit is characterized by the relative pitch of the key and termination syllables and the tone choice of the termination syllable. Brazil divided the pitch range of an utterance into three uniform dissections: low, mid, and high. The relative pitch of a prominent syllable was defined as the dissection in which its pitch resided. The tone choice of the termination syllable was specified by whether its pitch contour was rising, falling, level, rising then falling (rise-fall), or falling then rising (fall-rise).

Cambridge English Language Assessment (CELA) Corpus

The CELA corpus consists of 120 speech files of non-native English speaker's monologs from the speaking part of the Cambridge ESOL General English Examinations, which was previously used in Kang (2013). The speakers represented 21 first languages: 16 Spanish/Mexican, 11 Korean, eight Italian, seven Dutch, six French, five each of Chinese and Russian, four each of Greek, Portuguese, and Swedish, three German, two each of Swiss and Japanese, and one each of Arabic, Austrian, Bolivian, Brazilian, Bulgarian, Colombian, Estonian, and Turkish. Table 1 describes the Common European Framework of Reference for Languages (CEFR) proficiency level each of the speakers had attained, the equivalent Cambridge proficiency level, the number and gender of the speakers, and a description of the monologs they spoke.

Table 1

Cambridge English Language Assessment (CELA) Corpus

CEFR Proficiency Level	Cambridge Proficiency Level	Males	Females	Subject Of Monologues
B1	Preliminary English Test (PET)	16	16	The speaker is given a color photograph to discuss for one minute.
B2	First Certificate in English (FCE)	11	21	The speaker is provided with two photographs to talk about for one minute.
C1	Certificate in Advanced English (CAE)	11	23	The speaker selects two of three pictures and explains what is happening in the pictures for one minute.

C2	Certificate of Proficiency in English (CPE)	5	17	The speaker converses about a question from a card with various ideas on it for two minutes.

Automatic Scoring of English Speaking Proficiency of Unconstrained Speech

The English proficiency for a speaker is scored by the computer in three stages: (1) process the speech file to ascertain silent pauses, filled pauses, syllables, and the elements of Brazil's (1997) prosody model (i.e., tone units, prominent syllables, tone choices, and relative pitches); (2) compute 35 suprasegmental measures from the amounts and intervals of silent pauses, filled pauses, syllables, and the elements of Brazil's (1997) prosody model; and (3) utilize machine learning to analyze the suprasegmental measures and determine a proficiency score: B1, B2, C1, and C2. The following sections specify each of these stages.

Stage 1: Ascertain the Underlying Variables of the Suprasegmental Measures

A comprehensive discussion about ascertaining the underlying variables of the suprasegmental measures can be found in published articles (Johnson & Kang, 2015a; Johnson & Kang, 2015b) and manuscripts (e.g., Kang & Johnson, under review), which are currently under review for publication in other venues.

Stage 2: Compute the Suprasegmental Measures

Thirty-five suprasegmental measures shown in Table 2 are computed for each utterance based on the time intervals and amounts of silent pauses, filled pauses, syllables, and the four elements of Brazil's (1997) prosody model.

Table 2

Suprasegmental Measures

*Articulation rate	High-fall rate
Phonation time ratio	*Low-fall rate
Tone unit average length	*Mid-fall rate
*Syllable rate	*High-fall-rise rate
*Filled pause average duration	Low-fall-rise rate
Filled pause rate	Mid-fall-rise rate
Silent pause average duration	High-level rate
Silent pause rate	*Low-level rate
Prominent syllables per tone unit (i.e., pace)	Mid-level rate
*Percent of tone units with at least one prominent syllable	*High-rise-fall rate
Percent of syllables that are prominent (i.e., space)	Low-rise-fall rate
Overall pitch range	Mid-rise-fall rate
Non-prominent syllable average pitch	High-rise rate
Prominent syllable average pitch	*Low-rise rate
Paratone boundary onset pitch average height	*Mid-rise rate
Paratone boundary rate	Given lexical item mean pitch
Paratone boundary average pause duration	New lexical item mean pitch
Paratone boundary average termination pitch height	

The 35 suprasegmental measures were established from ones made use of in prior research (Brazil, 1997; Derwing, 1990; Derwing & Munro, 2001; Hincks, 2005; Kang et al., 2010; Kormos & Denes, 2004; Levis & Pickering, 2004; Pickering, 2004; Wennerstrom, 2001; Wichmann, 2000).

Stage 3: Utilize Machine Learning to Determine a Proficiency Score

In the final stage, a boosting ensemble of decision trees receives a subgroup of the suprasegmental measures (designated with an asterisk in Table 2) as input and outputs a proficiency score of B1, B2, C1, or C2. The boosting ensemble of decision trees was tested and trained using three-fold cross-validation of the 120 speech files. Each fold included 40 randomly allocated speakers, divided evenly by gender and proficiency.

The boosting ensemble did not utilize every one of the 35 suprasegmental measures to calculate a speaking proficiency score. The explanations for this are: (1) several of the original variables (i.e., quantities and time spans of silent pauses, filled pauses, syllables, and the four elements of the prosody model) that are utilized to compute the measures could be well correlated, and hence just one of them needs to be taken into account; (2) the measure might possibly not differ sufficiently across proficiency levels to be a suitable predictor; and (3) the original variables might contain inaccuracies, stemming from the intrinsic error rates of the equipment, procedures, and machine learning methods employed to ascertain them, which would make the suprasegmental measure an undependable proficiency prognosticator.

An exhaustive search for the best set of suprasegmental measures would necessitate an unfeasible assessment of 2.81×10^{40} permutations of the suprasegmental measures. To resolve this challenge, a genetic algorithm was utilized. A comprehensive discussion about the genetic algorithm can be found in manuscripts, which are currently under review for publication in other venues.

RESULTS

The objective of this research was to employ a collection of computer programs to automatically rate the oral proficiency of 120 speech files of non-native English examinee monologs from the speaking part of the Cambridge ESOL General English Examinations and to contrast the computer's ratings with the CELA examiners' ratings. The computer produced proficiency ratings of B1, B2, C1, and C2 utilizing the eleven suprasegmental measures shown in Table 3. The computer's proficiency ratings had a Pearson's correlation of 0.718 ($p < 0.01$) with the CELA examiner assigned proficiency ratings.

Table 3

Suprasegmental measures used by computer to rate unconstrained English speaking proficiency

Type	Suprasegmental Measure
Stress	Percent of tone units containing at least one prominent syllable
Pitch	Low-rise rate
	Mid-rise rate
	Low-level rate
	Low-fall rate
	Mid-fall rate
	High-rise-fall rate
	High-fall-rise rate
Pause	Filled pause average duration
Temporal	Syllable rate
	Articulation rate

DISCUSSION

Although not exactly the same as the CELA corpus, the English proficiency of speakers using unconstrained speech was scored automatically in four levels in two other studies. One of those

126

resulted in SpeechRater^{TM} which is described above (Zechner et al., 2009). In another study, Evanini and Wang (2013) used linear regression of ten features extracted from the output of an ASR configured to recognize the words to automatically score the spoken English responses given by non-native children in an English proficiency assessment of middle school students. The assessment included three different task types intended to measure a student's ability to converse in English. One of these, the Picture Narration task, is similar to the Cambridge test tasks. In the Picture Narration task, the child is presented with six pictures that portray a series of events and is asked to describe what is transpiring in the images. The Pearson's correlation between the scores assessed by the humans and those automatically scored was 0.62. This illustrates that the computer correlation of our method exceeds those of other similar computer programs (0.55-0.62). More importantly though, Zechner et al. (2009) reported human inter-rater reliability of 0.77 and Evanini and Wang (2013) reported 0.70 which also shows that the computer model for automatic scoring of unconstrained speech explained herein is nearing that of human raters with respect to inter-rater reliability.

CONCLUSION

In this paper, we presented a computer model for automatically scoring the English proficiency of unconstrained speech. In a test with the CELA corpus, the Pearson's correlation between the automatic scores from the computer model and the scores assigned by two human CELA examiners was 0.718. This correlation is greater than similar computer programs for automatically scoring the proficiency of unconstrained speech and is on the verge of inter-rater reliability of human scoring. The results also imply that stress, pitch, pause, and temporal suprasegmental measures might be the most important with regard to automated English proficiency scoring systems for unconstrained speech. This has also been shown to be true for human judgement (Kang et al., 2010).

Follow-on research that shows potential is expanding the computer model to automatically score the interactive aspects of English speaking proficiency. This bodes well for the reason that Brazil's (1997) model is markedly strong in elucidating the prosody of dialogs. Besides adding interactive measures to the computer model, augmenting the model with lexical and grammatical measures shows promise, too. A final area for further study is creating L1-specific models.

ABOUT THE AUTHORS

David O. Johnson is a post-doctoral researcher in the Applied Linguistics Speech Laboratory at Northern Arizona University, Flagstaff, AZ, USA. He is currently developing software and computer models to automatically rate English language proficiency. 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.

Okim Kang is an Associate Professor in the Applied Linguistics Program at Northern Arizona University. Her research concerns aspects of L2 pronunciation, speech perception and

production, automated speech scoring, oral language proficiency assessment, language attitudes, and World Englishes.

Romy Ghanem is a 3rd year PhD student in the Applied Linguistics program at Northern Arizona University, Flagstaff, AZ, USA. Her main research interests include speech perception (including nonnative speaker stereotyping and reverse stereotyping) and production, particularly accommodating different interlocutors' linguistic features. She has been working as a coder and trainer on a Speech Automated Recognition project for the past two years under the supervision of Okim Kang and David Johnson.

REFERENCES

Attali, Y., & Burstein, J. (2006). Automated essay scoring with e-rater® V. 2. The Journal of Technology, *Learning and Assessment*, 4(3).

Bernstein, J., Van Moere, A., & Cheng, J. (2010). Validating automated speaking tests. *Language Testing*.

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

Burstein, J., Kukich, K., Braden-Harder, L., Chodorow, M., Hua, S., Kaplan, B., Lu, C., Nolan, J., Rock, D., & Wolff, S. (1998). Computer analysis of essay content for automated score prediction: A prototype automated scoring system for GMAT analytical writing assessment essays. *ETS Research Report Series*, 1998(1), i-67.

Cambridge English Language Assessment (2015). www.cambridgeenglish.org, Retrieved March 29, 2015.

Chodorow, M., & Burstein, J. (2004). Beyond essay length: evaluating e-rater®'s performance on toefl® essays. *ETS Research Report Series*, 2004(1), i-38.

Chun, D. M. (2002). *Discourse intonation in L2: From theory and research to practice* (Vol. 1). John Benjamins Publishing.

Derwing, T. M. (1990). Speech rate is no simple matter. *Studies in Second Language Acquisition*, 12, 303– 313.

Derwing, T. M., & Munro, M. J. (2001). What speaking rates do non-native listeners prefer? *Applied Linguistics*, 22, 324–227.

Evanini, K., & Wang, X. (2013). Automated speech scoring for non-native middle school students with multiple task types. In *INTERSPEECH* (pp. 2435-2439).

Hincks, R. (2005). Measures and perceptions of liveliness in student oral presentation speech: A proposal for an automatic feedback mechanism. *System*, 33, 575–591.

Johnson, D. O., & Kang, O. (2015a). Automatic prominent syllable detection with machine learning classifiers. *International Journal of Speech Technology*, 18(4), 583-592.

Johnson, D. O., & Kang, O. (2015b). Automatic prosodic tone choice classification of Brazil's intonation model. *International Journal of Speech Technology*, DOI: 10.1007/s10772-015-9327-z.

Kang, O., & Rubin, D. L. (2009). Reverse linguistic stereotyping: Measuring the effect of listener expectations on speech evaluation. *Journal of Language and Social Psychology*.

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.

Kang, O. (2013). Linguistic analysis of speaking features distinguishing general English exams at CEFR levels B1 to C2 and examinee L1 backgrounds. *Research Notes, 52, 40-48*. http://www.cambridgeenglish.org/images/139525-research-notes-52-document.pdf

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

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological review*, 104(2), 211.

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

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

Rudner, L. M., Garcia, V., & Welch, C. (2006). An evaluation of IntelliMetric™ essay scoring system. *The Journal of Technology, Learning and Assessment*, 4(4).

Wennerstrom, A. (2001). *The music of everyday speech*. New York: Oxford University Press.

Wichmann, A. (2000). *Intonation in text and discourse*. London: Longman.

Zechner, K., Higgins, D., Xi, X., & Williamson, D. M. (2009). Automatic scoring of non-native spontaneous speech in tests of spoken English. *Speech Communication*, 51(10), 883-895.

Koffi. E. (2016). The acoustic phonetics of voiced the in seven varieties of L2-accented English: Focus on intelligibility. 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. 130-142). Ames, IA: Iowa State University.

THE ACOUSTIC PHONETICS OF VOICED TH IN SEVEN VARIETIES OF L2-ACCENTED ENGLISH: FOCUS ON INTELLIGIBILITY

Ettien Koffi, Saint Cloud State University, MN

The voiced interdental non-sibilant fricative [ð] produced by 10 native speakers of General American English (GAE) and 67 non-native speakers of English is analyzed quantitatively and acoustically. The quantitative data shows that GAE talkers produced [ð] accurately 88.09% of the time, substituted it with [d̠] 2.38%, and with [n̠] 9.52%. L2 talkers pronounced it accurately 47.51% of the time, substituted it with [d̠] 44.27%, with [z] 4.22%, with [s] 1.74%%, with [ɾ] 1.4%, and with [θ] .99%. A closer scrutiny of the distributional patterns reveals that all the substitutions occurred only in syllable onsets, but not intervocalically. An acoustic phonetic analysis indicates that [d̠] is by far the preferred substitute for [ð] because their intensities are perceptually indistinguishable. A relative functional load (RFL) analysis is done also to gauge the likelihood of unintelligibility if/when [ð] is replaced by [d̠], [z], or [v].

INTRODUCTION

The L2 phonology literature is replete with two hyperbolic claims about the pronunciation of voiced TH. The first gives the impression that native speakers of GAE consistently and predictably produce it as [ð]. The second assumes a priori and on the basis of piecemeal evidence that non-native speakers cannot produce it accurately (Jenkins 2000, pp. 137-138, Celce-Murcia et al. 2010, p.82, Lado and Fries 1954, p. 97-102, Prator and Robinett 1985, pp. 148-150, to mention only these). Quantitative and instrumental analyses of 444 occurrences of [ð] are done to see if these claims are supported linguistically.

DISTRIBUTION OF [ð] IN WORLD LANGUAGES

To start with, let's familiarize ourselves with the distribution of [ð] in world languages. Maddieson (1984, pp. 43-45) found that [ð] occurred in only 21 (6.62%) of the 317 languages surveyed in UPSID (the UCLA Phonological Segment Inventory Database). In other words, [ð] is an uncommon sound. Table 1 shows that only two languages of the seven considered in this paper have [ð] in their phonetic inventory. In Spanish, it occurs as an allophone of /d/. It is also found in some dialects of Arabic.

The lack of [ð] in the languages under investigation raises the following questions:

1. Can L2 speakers of English whose native language(s) lack [ð] manage to produce it accurately?
2. If they cannot, what segments do they use to substitute for it?
3. Does the compensatory strategy used interfere with segmental intelligibility?

Table 1

Distribution of [ð]

Languages		Place			
		Labio-dental	Interdental	Alveolar	Palatal
English	**Manner**	v	ð	z	ʒ
Arabic		__	ð	z	__
Japanese		β	__	z	x
Korean		__	__	z	__
Mandarin		__	__	ʐ	__
Slavic		ʋ	__	z	ʒ
Somali		__	__	__	__
Spanish		β	ð	__	__

PREDICTED AND OBSERVED SUBSTITUTIONS

Hatten (2009) provides the data in Table 2 that helps answer the first two questions:

Table 2

Substitutes of [ð]

	Languages	**[v]**	**[z]**	**[d̪]**
1.	Afrikaans	☐		
2.	Brazilian Portuguese	☐		
3.	Canadian French			☐
4.	Cantonese		☐	
5.	Czech			☐
6.	Dutch			☐
7.	Egyptian Arabic		☐	
8.	French		☐	
9.	German		☐	
10.	Hebrew			☐
11.	Italian			☐
12.	Japanese		☐	
13.	Korean			☐
14.	Norwegian			☐
15.	Polish			☐
16.	Russian			☐
17.	Spanish			☐
18.	Swedish			☐
19.	Yiddish			☐
	Total	2	5	12

We see that [d̪] is by far the most common substitute for [ð]. The segment [z] is a distant second, followed further back by [v]. It must be noted that Hatten obtained his data by

soliciting opinions on Linguistlist.org. For this reason, his findings must be taken with a grain of salt.

METHODOLOGY

The methodology used in this study is based on the actual pronunciation of voiced TH obtained from 77 participants who read the following text: [1]

> *Please call Stella. Ask her to bring **these** things with her from **the** store: Six spoons of fresh snow peas, five thick slabs of blue cheese, and maybe a snack for her **brother** Bob. We also need a small plastic snake and a big toy frog for **the** kids. She can scoop **these** things into three red bags, and we will go meet her Wednesday at **the** train station.*

The text contains six occurrences of [ð]: <the> x3, <things> x2, <brother> x1. The quantitative analysis is based on the impressionistic IPA transcriptions of [ð] made by trained phoneticians from George Mason University (GMU). The participants include seven native GAE talkers who attempted 42 [ð]s (7 x 6) and 67 L2 speakers of English (10 Arabic, 10 Japanese, 10 Korean, 10 Mandarin, 11 Slavic, 6 Somali, and 10 Hispanic talkers) who attempted [ð] 402 times (67 x 6). Figure 1 breaks down the percentages of accuracy as follows:
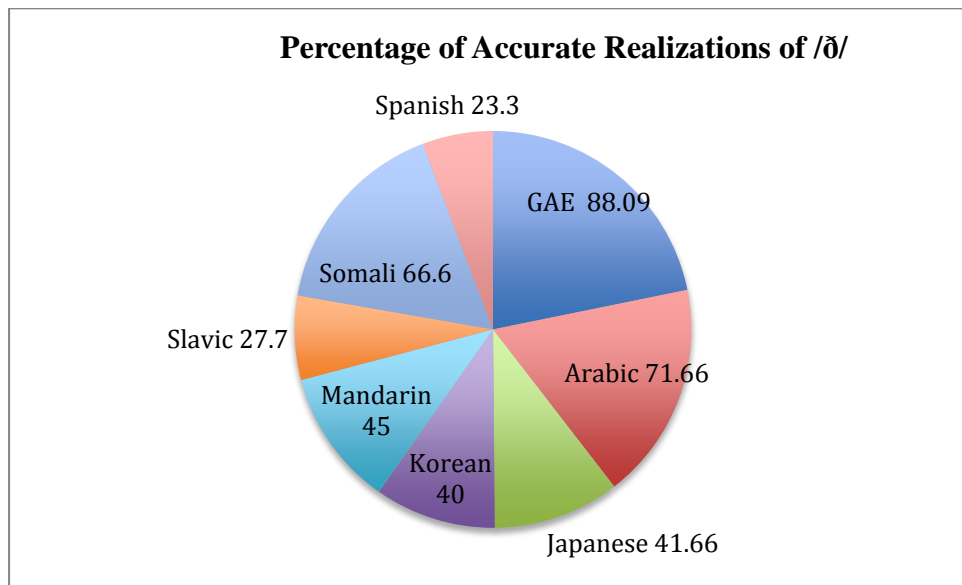
Figure 1. Production accuracy of [ð]

Three cursory observations are in order: first, not all native speakers produced [ð] accurately (11.91%). Secondly, many non-native speakers produced [ð] accurately

---

[1] The text is found at: http://accent.gmu.edu/browse.php. A more detailed methodology is found in Koffi (2015), a prequel article, available at: http://repository.stcloudstate.edu/stcloud_ling/vol4/iss1/2/.

(44.27%). Thirdly, the rate of accuracy in the production of [ð] varies by language group. Figure 2 highlights the segments used to substitute for [ð] and their frequencies:
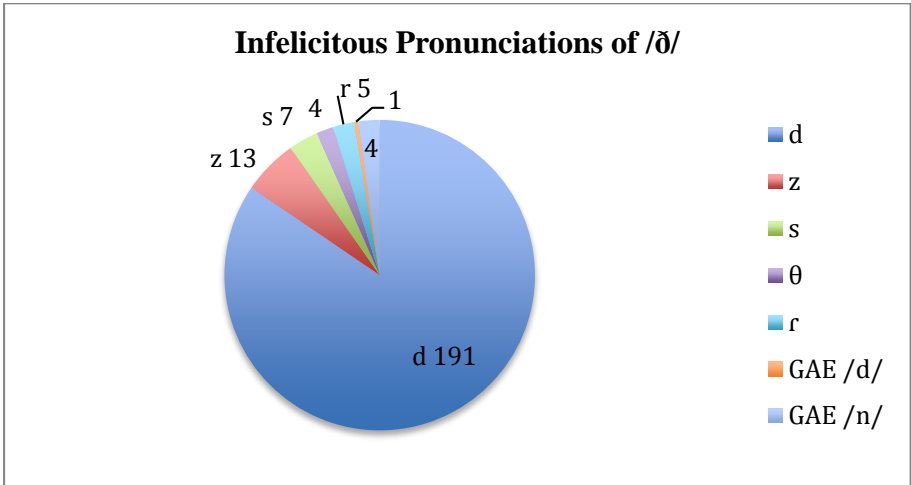
Figure 2. Segments substituted for [ð].

The impressionistic transcription data shows clearly that [d̪] is the preferred substitute for [ð], followed by [z], which is a distant second. It is worth noting that there is not a single case in the GMU transcriptions where [ð] is replaced by [v]. Miscellaneous substitutions include [ɾ], [s], and [θ]. They will not be addressed in this paper because I consider them to be slips of the tongue given their sporadic nature.

Variable Pronunciations of [ð] by GAE Talkers

The spectrogram in Figure 3 shows that a GAE talker, TN 23M, pronounces the TH of <these> in the phrase <these things> as a [d̪].
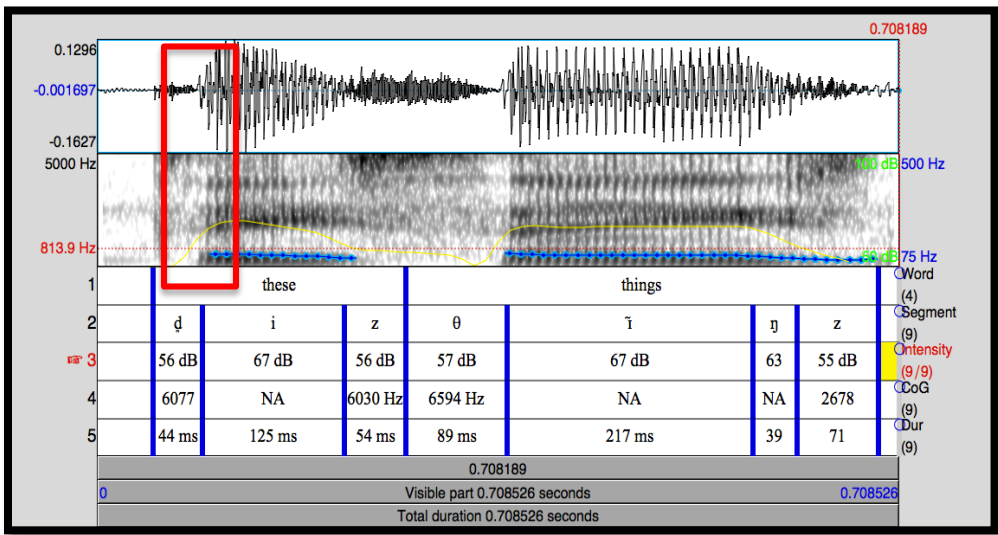
Figure 3. Spectrogram of [ð] Produced as [d̪].

The highlighted portion shows a weakly dentalized [d̪]. The lack of frication noise indicates that TN 23M does not produce the voiced TH as a fricative, but rather as a stop. There is mounting sociophonetic evidence that this pronunciation is widespread among GAE talkers. Several contributors to *American Voices: How Dialects Differ from Coast to Coast (2006)* have made comments in this regard. Newman (2006, p.85) has found this pronunciation in New York City. Simon (2006, p.132) writes that it is an unmistakable feature of the dialect of English spoken in Michigan's Upper Peninsula, and in Milwaukee. Frazer (2006, p.104) shows that it occurs in the speech of Finnish-Americans, Swedish-Americans, and Norwegian-Americans in Minnesota and Wisconsin. Eckert (2008, pp. 27-28) lists German-Americans, Cajuns, Polish-Americans, and Chicano speakers among the ethnic groups that pronounce [ð] as [d̪]. Zsiga (2013, p. 439) adds speakers of African American English to this list. In a nutshell, millions of GAE talkers routinely substitute [d̪] for [ð].

The sound [n̪] also appears as an allophone of [ð]. KY 89M and MN 143M each produce it twice as [n̪]. These realizations occur in <bring **th**ese things> and <from **th**e store>. The conditioning environment for this pronunciation can be stated formally as follows:

$$/ð/ \rightarrow [n̪] \: / \: [+cons, +nasal] \: \_\_\_\_$$

Ogden (2009, p.128) notes that "For many speakers, [ð] in particular is highly variable in its manner of articulation, ranging through plosive, *nasal*, fricative, lateral approximant, and approximant articulations." The word "nasal" has been italicized for emphasis. The nasalization of [ð] is particularly noticeable in prepositional phrases such as <in the …>, <on the …>, <from the …> or with conjunctions in phrases such as <when the …>, <then the …>, or <than the …>.

Variable Pronunciations of [ð] in L2 Englishes

The 67 non-native speakers of English attempted a total of 402 [ð]s. They produced it accurately as [ð] 178 times. This also means that they mispronounced it 224 times (55.72%). In such instances, [ð] was overwhelmingly replaced by [d̪] 191 times (47.51%), and to a much lesser extent by [z] 13 times (3.23%). An instrumental analysis was performed on the various attempts of [ð] to see if there is any acoustic rationale for these substitutions. Various correlates were examined, but the one that is most relevant for this paper is intensity. Ladefoged and Maddieson (1996, p.139) contend that it is by far the most robust correlate for the study of fricatives.

Table 4

The Intensities of [ð]and its Substitutes

	five	please	these	Wednesday
Languages	v	z	ð	d̪
Arabic mean	66.31	68.19	68.36	70.70
Japanese mean	68.80	67.60	65.35	65.19
Korean mean	65.68	71.24	65.46	61.46
Mandarin mean	64.35	65.63	66.97	65.97
Slavic mean	60.55	63.54	67.10	61.68
Somali mean	62.31	67.43	68.17	63.14
Spanish mean	65.95	68.64	65.13	63.35
GAE mean	61.16	70.96	67.10	67.72

A comparison of the means and relative intensities shows that the acoustic distance between [ð] and [d̪] on the one hand, and [ð] and [z] on the other, is less than 3 dB in most cases. Decades of experimental acoustic research have shown that when two sounds differ in intensity by 3 dB, they are barely distinguishable to the naked ear. (Hasen 2001, p.41).   For two sounds to be perceived as aurally distinct, their relative intensity distance must be ≥ 5dB. GAE talkers who substitute [d̪] (67.72 dB) for [ð] (67.10 dB) do so because both sounds are indistinguishable acoustically. Polka et al.'s study (2001:2193) confirms my findings. They found that [ð] (56.2 dB) and [d̪] (53.7 dB) were often substituted for each other.   This can be explained by the fact that the relative intensity distance between them is below the 3 dB threshold.   The Fricative Intelligibility Criterion (FIC) can help to explain why [d̪], [z], and to a lesser extent [v] are often substituted for [ð]:

Fricative Intelligibility Criterion (FIC)

> A segment that is acoustically closely related to [ð] can be substituted for it without interfering with intelligibility if and only if that segment agrees with it in place of articulation, in voicing, and if the intensity distance between the two is ≤ 3 dB.

Phonetically speaking, [d̪], [z] and [v] are legitimate substitutes for [ð] because they meet all three criteria in FIC.

Intelligibility Assessment

We are now is a position to answer the third research question, namely, does replacing [ð] with [d̪], [z], or [v] interfere with intelligibility?   This question is answered for each individual segment by relying on Catford's (1987, pp. 87-100) relative functional load (RFL) data.

Intelligibility of [ð] Pronounced as [d̪]

The segment [d̪] is substituted for [ð] more often than any other segments (191 times out of 402 occurrence, i.e., 47.51%). The segment [d̪] is a legitimate candidate because in all but Slavic and Somali-accented Englishes, it has perceptually the same intensity as [ð]. This suggests that L2 talkers derive their pronunciation cues from the intensity of the segment they hear. Another explanation for why [d̪] is the preferred substitute for [ð] may have something to do with the fact that it is widespread cross-linguistically. Maddieson (1984, p. 32, Table 2.5) notes that it is found in 316 of the 317 languages in the UPSID   Furthermore, substituting [d̪] for [ð] interferes minimally, if at all, with intelligibility. The relative functional load between them is 19% in word-initial position, and 5% at the end of words. Some words such as <they> vs. <day>, <then>, vs. <Dan>, <then> vs. <den>, <those> vs. <doze>, and <though> vs. <dough> that may cause misunderstandings (Higgins 2010). [2] However, the syntactic context in which they occur can reduce unintelligibility.

Intelligibility of [ð] Pronounced as [z]

The segment [z] is used to replace [ð] only 13 times (3.23%). Eight of the substitutions are made by Japanese 4M (three times), Japanese 12M (three times), and Japanese 13M (two times). The remaining five are by Mandarin 9M alone. Replacing [ð] with [z] does interfere with intelligibility at the beginning of words because the two segments are not contrastive in this position. Furthermore, the RFL between them is only 1% at the end of words. Words that can cause misunderstanding are <clothe> vs. <cloze>, <breathe> vs. <breeze>, <seethe> vs. <seize>, <teethe> vs. <teas>, <teethe> vs. <tease>, <tithe> vs. <ties>.   However, in all these instances, the syntactic context can help minimize unintelligibility. Even so, Jenkins' (2000:138) notes that this is a stigmatized substitution that "elicit responses of irritation when used with high frequency." Francophone learners of English should heed this admonition because they are prone to pronouncing most of their [ð]s as [z]s (Fromkin et al. 2014:2, Prator and Robinett 1985, p. 149).

Intelligibility of [ð] Pronounced as [v]

Table 2 lists Brazilian Portuguese as one of only two languages out 19 that substitute [v] for [ð]. When I presented this data to three Brazilian applied linguists who used to teach English as a Foreign Language in Brazil, they objected. They all stated without hesitation that their students and acquaintances routinely replace [ð] with [d̪]. Regardless, it is clear from the GMU data and from Hatten (2009) that [v] is not commonly used to substitute for [ð].   This calls into question the merits of Jenkins' (2000, p. 138) proposal that [v] be taught to L2 learners of speakers of English whose native languages lack [ð]. Her proposal is problematic for three reasons. First, it lacks strong empirical support. She made this proposal on the basis of piecemeal evidence from speech samples obtained from six L2 speakers representing only three language groups (Jenkins 2000, pp. 59-61, 84). Secondly, since [v] is found in only 21.13% of the languages in UPSID, what should a teacher do if the student's L2 lacks [v]? Thirdly,

---

[2] http://myweb.tiscali.co.uk/wordscape/wordlist/m<advinimal.html. Retrieved on October 2nd 2015.

aerodynamic factors explain why [v] is relatively infrequent in world languages. Johnson (2012, p. 156) remarks that it is articulatorily challenging to produce. These misgivings notwithstanding, if a speaker substitutes [v] for [ð], intelligibility is not likely to be compromised because the RFL between them is 11% in word initial positions, and 1% at the end of words. The only lexical minimal pairs or near minimal pairs that could cause intelligibility problems are words such as <further> vs. <fervor>, <clothes> vs. <cloves>, <loathes> vs. <loaves>, <slither> vs. <sliver>, <than> vs. <van>, and <then> vs. <van>. Here too the syntactic context can alleviate possible misunderstandings.

Pedagogical Implications

The impressionistic transcription data shows that all 67 attempts of [ð] in <brother> were successful irrespective of the talker's language background or level of proficiency. This is a significant finding of great pedagogical importance. It means that the difficulties associated with producing [ð] accurately are limited only to syllable onsets, but not to the intervocalic environment where it is produced with 100% accuracy. This finding has both curricular implications. L2 learners would do well if [ð] is first taught to them when it occurs between vowels. There are a dozen or so high frequency words such as <bother>, <breather>, <brother>, <either>, <farther>, <father>, <further>, <mother>, <other>[3], <rather>, <worthy> and <without> in which [ð] occurs intervocalically. This list can be augmented with words such as <breathe>, <bathe>, <loathe>, <seethe>, <teethe>, <tithe>, and <writhe> in which the silent <e> gives the impression that [ð] occurs between vowels. Once students have mastered the pronunciation of [ð] in this environment, the teacher can introduce it in syllable onsets.

SUMMARY

The preceding analyses have shown that voiced TH has variable pronunciations in both L1 and L2 English. Many GAE talkers produce it in three ways: some as [ð], some as [d̪], and others as [n̪]. The variant [d̪] is ethnolinguistically conditioned, whereas the variant [n̪] is phonologically conditioned. Most L2 speakers of English in this study replace [ð] with [d̪], and to a lesser extent with [z] when it occurs in syllable onsets. However, between two vowels, they produce it accurately 100% of the time. The choice of [d̪] or [z] as substitutes for [ð] is justified acoustically and cross-linguistically. These two segments have relative intensities that fall within the ≤ 3 dB threshold. This makes them practically aurally indistinguishable from [ð]. Cross-linguistically, [d̪] is best suited as a substitute for [ð] because it is found in more than 99% of the languages in UPSID. The substitutions do not compromise intelligibility because the relative functional loads between [ð], [d̪], and [z] are very low, even negligible. None of the 67 L2 talkers in this study replaces [ð] with [v]. There are two reasons for this. First, it is a relatively uncommon sound. Secondly, its aerodynamic properties make it hard to produce. Consequently, it should not be seen as a legitimate substitute for [ð].

---

[3] Jenkins (2000, p.59) claims in Extract 2 that the talker pronounced <other> as [ɔd̪əɹ]. This transcription should be regarded as highly suspicious because she indicates in parentheses that it is unintelligible.

ABOUT THE AUTHOR

Ettien Koffi is a professor of Linguistics. He teaches the linguistics courses in the TESOL/Applied Linguistics MA program in the English Department at Saint Cloud State University, MN. He has written three linguistic books: *Language Society in Biblical Times* (1996), *Paradigm Shift in Language Planning and Policy: Game Theoretic Solutions* (2012), and *Applied English Syntax* (2015). He is the author of many peer-reviewed articles on various topics in linguistics. His primary area of specialization is at the interface between acoustic phonetics and phonology. He has extensive experience in emergent orthographies and in the acoustic phonetic and phonological description of dialect variation. He can be reached via email at: *enkoffi@stcloudstate.edu.*

REFERENCES

Catford, J. C. (1987). Phonetics and the Teaching of Pronunciation. In J. Morley (Ed.), *Current Perspectives on Pronunciation: Practices Anchored in Theory*. Washington, DC: Teachers of English to Speakers of Other Languages.

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

Eckert, P. (2008). Where do Ethnolects Stop? *International Journal of Bilingualism, 12* (1), 25-42.

Frazer, T. C. (2006). An Introduction to the Midwest English. *American Voices: How Dialects Differ from Coast to Coast*, ed. by Walter Wolfram and B. Ward, pp. 101-105. Malden, MA: Blackwell Publishing.

Fromkin, Victoria, Robert Rodman; and Nina Hyams. (2014). *An Introduction to Language, 10th ed.*, New York: Cengage Learning.

Hatten, David J. (2009). Substitution Patterns for English Inter-dental Fricatives by L1 Latin American Spanish-Speakers. Retrieved from: http://www.academia.edu/13539048/Substitution_Patterns_for_English_Interdental_Fricatives_by_L1_Latin on September 26, 2015. Ladefoged, P. & Keith K. (2015). *A course in phonetics*, 7th edition. Stamford, CT: Cengage Learning.

Higgins, J. 2010. Minimal Pair for English RP: Lists by John Higgins. http://myweb.tiscali.co.uk/wordscape/wordlist/m<advinimal.html. Retrieved on October 2nd, 2015.

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

Johnson, K. (2012). *Acoustic and Auditory Phonetics, 3rd ed.,* Malden, MA: Wiley-Blackwell.

Koffi, E. (2015). The Pronunciation of Voiceless <TH> in Seven Varieties of L2
        Englishes: Focus on Intelligibility. *Linguistic Portfolios*, 4: 2-23. Available
        online at: http://repository.stcloudstate.edu/stcloud_ling/vol4/iss1/2/.

Lado, R. & Fries C. C. (1962). *English Pronunciation: Exercises in Sound Segments,
        Intonation, and Rhythm*.   Ann Arbor: The University of Michigan Press.

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

Newman, M. (2006). New York Tawk (New York City, NY. *American Voices: How
        Dialects Differ from Coast to Coast*, ed. by Walter Wolfram and B. Ward, pp. 82-
        87. Malden, MA: Blackwell Publishing.

Ogden, R. (2009). *An Introduction to English Phonetics*. Edinburg, Scotland: Edinburgh
        University Press Ltd.

Prator, Jr. C. & Robinett, B. (1985). *Manual of American English Pronunciation, 4th
        edition.* New York: Harcourt Brace & Company.

Zsiga, E.C. (2013). *The Sounds of Language: An Introduction to Phonetics and
        Phonology*. Malden, MA: Wiley-Blackwell.

APPENDIX
Inventory for [ð]

	These	The	Brother	the	These	The	Total [ð]	Total [d]	Total [v]	Total [n]	Total [ ]
KY150	[ð]	[ð]	[ð]	[ð]	[ð]	[ð]	6				
KY89	[n]	[n]	[ð]	[ð]	[ð]	[ð]	4			2	
CA32	[ð]	[ð]	[ð]	[ð]	[ð]	[ð]	6				
GA330	NA	NA	NA	NA	NA	NA					
MN143	[n]	[n]	[ð]	[ð]	[ð]	[ð]	4			2	
NY6	[ð]	[ð]	[ð]	[ð]	[ð]	[ð]	6				
OR184	NA	NA	NA	NA	NA	NA					
TN23	[ð]	[ð]	[ð]	[ð]	[ð]	[d]	5	1			
TX286	NA	NA	NA	NA	NA	NA					
VA16	[ð]	[ð]	[ð]	[ð]	[ð]	[ð]	6				
All total	42						37	1		4	

Total of expected [ð] = 42; total [ð] produced as [ð] 37 (88.09%); total [ð] as [d] 1 (2.38%) total [ð] produced as [n] = 4 (9.52%)

	These	The	Brother	The	These	The	Total [ð]	Total [d]	Total [s]	Total [z]	Total [θ/f]
Arabic 1F	[ð]	[ð]	[ð]	[ð]	[ð]	[ð]	6				
Arabic 30F	[d]	[ð]	[ð]	[ð]	[ð]	[ð]	5	1			
Arabic 35M	[d]	[ð]	[ð]	[ð]	[ð]	[ð]	5	1			
Arabic 36M	[d]	[ð]	[ð]	[ð]	[ð]	[ð]	5	1			
Arabic 40M	[ð]	[ð]	[ð]	[ð]	[ð]	[ð]	6				
Arabic 44F	[d]	[d]	[ð]	[ð]	[d]	[d]	2	4			
Arabic 46M	[d]	[d]	[ð]	[ð]	[d]	[d]	2	4			
Arabic 47M	[d]	[ð]	[ð]	[ð]	[ð]	[ð]	5	1			
Arabic 50M	[d]	[d]	[ð]	[d]	[ð]	[d]	4	2			
Arabic 51M	[ð]	[d]	[ð]	[d]	[d]	[ð]	3	3			

Total of expected [ð] = 60; total [ð] produced as [ð] 43 (71.66%); total [ð] produced as [d]: 17 (28.33%).

	These	The	Brother	the	These	The	Total [ð]	Total [d]	Total [s]	Total [z]	Total [θ]
Japanese 2F	[d]	[d]	[ð]	[d]	[ð]	[ð]	3	3			
Japanese 3F	[ð]	[d]	[d]	[ð]	[d]	[d]	2	4			
Japanese 4M	[d]	[z]	[z]	[ð]	[d]	[z]	1	2		3	
Japanese 5F	[ð]	[ð]	[ð]	[d]	[ð]	[ð]	5	1			
Japanese 8M	[d]	[ð]	[ð]	[ð]	[ð]	[ð]	5	1			
Japanese 9M	[θ]	[ð]	[ð]	[ð]	[ð]	[ð]	5				1
Japanese 10F	[d]	[d]	[ð]	[ð]	[d]	[d]	2	4			
Japanese 11F	[d]	[d]	[d]	[d]	[d]	[d]		6			
Japanese 12M	[d]	[z]	[z]	[z]	[d]	[s]		2	1	3	
Japanese 13M	[d]	[ð]	[ð]	[z]	[d]	[z]	2	2		2	
All total							25	25	1	8	1

Total of expected [ð] = 60; total [ð] produced as [ð] 25 (41.66%); total [ð] as [d] 25 (41.66%) total [ð] produced as [s] = 1 (1.66%); total [ð] produced as [z] = 8 (13.33%); total [ð] produced as [θ] = 1 (1.66%)

	These	The	Brother	the	These	The	Total [ð]	Total [d]	Total [v]	Total [f]	Total [θ]
Korean 1M	[d]	[d]	[ð]	[d]	[d]	[ð]	2	4			
Korean 2F	[ð]	[d]	[ð]	[d]	[d]	[ð]	3	3			
Korean 3F	[ð]	[d]	[ð]	[ð]	[d]	[d]	3	3			
Korean 4F	[d]	[d]	[ð]	[ð]	[d]	[ð]	3	3			
Korean 8F	[d]	[d]	[ð]	[θ]	[f]	[d]	1	3			2
Korean 9M	[ð]	[ð]	[ð]	[d]	[θ]	[d]	3	2			1
Korean 10M	[d]	[d]	[ð]	[d]	[d]	[d]	1	5			
Korean 11M	[d]	[d]	[ð]	[ð]	[ð]	[d]	3	3			
Korean 12M	[d]	[d]	[d]	[ð]	[d]	[d]	1	5			
Korean13F	[ð]	[d]	[ð]	[ð]	[ð]	[d]	4	2			
All total							24	33			3

Total of expected [ð] = 60; total [ð] produced as [ð] 24 (40%); total [ð] as [d] 33 (55%);
total [ð] produced as [θ] = 3 (5%)

	These	The	Brother	the	These	The	Total [ð]	Total [d]	Total [z]	Total [ɾ]	Total [ʔ]
Mandarin1F	[ð]	[ð]	[ð]	[ð]	[ð]	[ð]	6				
Mandarin2F	[d]	[d]	[ð]	[ð]	[d]	[ð]	3	3			
Mandarin3M	[d]	[ð]	[ð]	[ð]	[ð]	[ɾ]	5			1	
Mandarin4F	d	d	d	d	[ð]	d	1	5			
Mandarin5F	d	?	d	[ð]	d	?	1	3			2
Mandarin6F	[ð]	[d]	[ð]	[ð]	[d]	[ð]	4	2			
Mandarin8M	d	n ?	ɾ	d	d	d	4			1	1
Mandarin9M	z	z	[ð]	z	z	z	1		5		
Mandarin12M	d	d	ɾ	[ð]	d	d	1	4		1	1
Mandarin19M	d	d	d	[ð]	d	d	1	5			
All total							27	21	5	3	4

Total of expected [ð] = 60; total [ð] produced as [ð] 27 (45%); total [ð] as [d] 21 (35%)
total [ð] produced as [z] = 5 (8.3%); total [ð] produced as "others" = 7 (11.66%)

	These	The	Brother	The	These	The	Total [ð]	Total [d]	Total [v]	Total [ɾ]	Total [ ]
Croatian1F	[d]	[ð]	[ð]	[d]	[d]	[d]	2	4			
Croation2F	[d]	[d]	[d]	[d]	[d]	[ɾ]		5		1	
Croatian4M	[d]	[d]	[ð]	[d]	[d]	[ɾ]	1	4		1	
Croatian5F	[d]	[d]	[d]	[d]	[d]	[d]		6			
Croation6F	[d]	[d]	[ð]	[d]	[d]	[d]	1	5			
Serbian1F	[d]	[t]	[d]	[d]	[d]	[t]		4		2	
Serbian2M	[d]	[d]	[ð]	[ð]	[d]	[ð]	3	3			
Serbian6M	[d]	[d]	[ð]	[d]	[d]	[d]	1	5			
Serbian11M	[ð]	[ð]	[ð]	[d]	[ð]	[t]	4	1		1	
Serbian12F	[d]	[d]	[ɾ]	[d]	[d]	[d]	5				1
Serbian14F	[d]	[d]	[ð]	[d]	[d]	[d]	1	5			
All total							18	42		5	1

Total of expected [ð] = 66; total [ð] produced as [ð] 18 (27.7%); total [ð] as [d] 42
(63.3%); total [ð] produced as "others" = 6 (9.9%)

	These	The	Brother	the	These	The	Total [ð]	Total [d]	Total [v]	Total [θ]	Total [0]
Somali 1F	[d]	[ð]	[ð]	[ð]	[d]	0	3	2			1
Somali 2M	[ð]	[d]	[ð]	[d]	[d]	[d]	2	4			
Somali 3F	[d]	[d]	[ð]	[ð]	[d]	[d]	2	4			
Somali 4F	[ð]	[ð]	[ð]	[ð]	[ð]	[ð]	6				
Somali 5M	[ð]	[ð]	[ð]	[ð]	[ð]	[ð]	6				
Somali 6M	[ð]	[ð]	[ð]	[ð]	[θ]	[ð]	5			1	
All total							24	10		1	1

Total of expected [ð] = 36; total [ð] produced as [ð] 24 (66.6%); total [ð] as [d] 10 (27.7%) total [ð] produced as [θ] = 1 (2.7%)

	These	The	brother	the	These	The	Total [ð]	Total [d]	Total [ɾ]	Total [v]	Total [ ]
Spanish 1M	d	d	[ð]	d	[ð]	d	2	4			
Spanish 2M	d	[ð]	ð	d	[ð]	d	2	4			
Spanish 4F	d	d	ð	d	[ð]	d	2	4			
Spanish 6M	d	d	ð	[ð]	[ð]	d	3	3			
Spanish 9M	d	d	ð	[ð]	d	d	2	4			
Spanish 11F	d	[ð]	ð	d	d	d	1	5			
Spanish 13F	d	d	ð	d	[ð]	d	1	5			
Spanish 14F	d	d	ð	ɾ	[ð]	d		5	1		
Spanish 16M	d	d	ð	[ð]	[ð]	d	3	3			
Spanish 20M	d	d	ð	d	d	d	1	5			
All total							14	45	1		

Total of expected [ð] = 60; total [ð] produced as [ð] 14 (23.3%); total [ð] as [d] 45 (75%); total [ð] produced as "others" = 1 (1.66%)

AN EXPLORATION OF TEACHING INTONATION USING A TED TALK

Alison McGregor, University of Texas at Austin
Beth Zielinski, Macquarie University
Colleen Meyers, University of Minnesota
Marnie Reed, Boston University

With the intent of researching the phenomenon of American English intonation to inform classroom practices, this study investigated the intonation of a TED Talk in a multi-layered approach – interpretatively, perceptually and acoustically (Vaissière, 2005). To identify the features produced in the monologic speech sample of a speaker of North American English, the analysis included 1) interpretative judgments of the speaker's syntactic and information structure and attitude; 2) perceptual judgments of thought groups, key, prominence, salience and tone choice and 3) acoustic analysis of pauses and features identified in the perceptual analysis. PRAAT was used for the acoustic analysis of pauses, pitch range, and prominence (Levis & Pickering, 2004). Triangulated results show an integrated view of language and intonation highlighting relationships between component aspects of intonation and among the interpretative, perceptual, and acoustic layers. Findings support the contextualized teaching of intonation as well as pedagogical use of rich TED Talk speech samples. Findings also indicate a pedagogical shift from teaching functions of intonation in isolation toward an integrative approach, which embraces the overlapping layers that create meaning-making.

INTRODUCTION

Over a decade ago, Levis (1999) argued that the treatment of intonation in textbooks had remained the same for 30-50 years due to materials being based on an inadequate view of intonation and a focus on decontextualized sentence-level practice. A lack of connection between intonation research, teaching, and textbook inclusion may have stemmed in part from the historically different views by British and American scholars in terms of the definition, assumptions, terminology, purpose (describe versus prescribe), methods and approaches (Levis, 2005). Despite contemporary advances in the field, however, Reed and Michaud (2015) still observe some 15 years later that, "intonation remains a challenge for teachers and students alike" (p. 454). Paunović and Savić (2008) argue that "teachers do not seem to be theoretically or practically well‑equipped to explain and illustrate its significance" (pp. 72-73) and it stands to reason that teacher cognition (Baker & Murphy, 2011), a teacher's belief and knowledge about intonation teaching and learning, will directly impact classroom practices revolving around intonation. Teacher training textbooks also undoubtedly influence the conundrum. One of the most popular teacher training books *Teaching Pronunciation* (Celce-Murcia, Brinton, and

Goodwin, 2010), for example, is rich in information yet represents intonation to teachers as a descriptive list of mechanical parts and functions.

Consequently, two significant gaps exist in (i) combining different approaches to the research on intonation (namely, interpretative, perceptual, and acoustic approaches) and (ii) applied research to inform classroom practices in the teaching and learning of intonation in a practical way. To fill these gaps, this study employed a multi-layered integrative approach to the analysis and description of discourse-level intonation to inform classroom practice.

This study focused on intonation because it contributes significantly to a speaker's communicative effectiveness; that is, the extent to which s/he engages with the audience in a range of speaking contexts (Hincks, 2005; Hincks & Edlund, 2009; Pickering, 2001, 2004; Slater, Levis & Muller Levis, 2015). Like Levis and Wichmann (2015), we define intonation as "the use of pitch variations in the voice to communicate phrasing and discourse meaning in varied linguistic environments" (p. 137). We focused on TED Talks because they provide readily available models of effective speakers and serve as a rich, authentic, and contextualized resource for ESL classrooms. Furthermore, publishers are now utilizing TED Talks and incorporating them into their English language teaching materials (e.g., World English series published by National Geographic Learning and Cengage Learning; see for example, Chase, Milner, & Johannsen, 2015). However, as yet there has been no systematic integration of pronunciation related to TED Talks into these texts. It is therefore difficult for teachers to capitalize on the potential teaching resource that TED Talks and other online materials offer in regard to the teaching of intonation.

The aim of the study was to use a 3-layered framework (interpretative, perceptual, acoustic levels of analysis) proposed by Vaissière (2005) to analyze the intonation of a TED Talk by a native North American, Matt Cutts (Cutts, 2011). The following research questions were addressed.

> 1. How does an interpretative, perceptual, and acoustic analysis of Matt Cutts' intonation elucidate the perceived impact of his TED Talk?
>
> 2. How do the results of an interpretative, perceptual, and acoustic analysis of intonation inform teaching practices?

METHODS

Segmentation and Trial Analysis

We chose the TED Talk by Matt Cutts (Cutts, 2011) because we felt it provided a good example of an inspiring, persuasive talk delivered by a native North American speaker. For the analysis, we used the online transcript of the talk provided by TED with paragraphs and grammatical marks (periods and commas) removed, and used verbal pauses to identify and mark utterances (//) and thought groups (/). Pitch range (key) and topic changes were also used as indicators of the speech paragraphs. This generated a working script for the interpretative and perceptual analysis (see Appendix A for marked-up transcript).

For calibration and reliability purposes, we performed a trial analysis and moderation on the first speech paragraph of the TED Talk (Cutts, 2011, 0:12 - 0:42). This trial employed the 3-layered framework proposed by Vaissière (2005), and was conducted by three researchers (two American and one Australian) performing independent interpretative and then perceptual analyses, followed by collation and moderation. The final step of the trial analysis was an acoustic analysis using the software program PRAAT.

Coding and Analysis

We analyzed the second speech paragraph of the TED Talk (Cutts, 2011, 0:43 - 1:27) at the interpretative, perceptual, and then acoustic levels. Elements of each level of analysis are presented in Tables 1, 2, and 3 respectively.

Table 1

Interpretative Analysis Components

Functions of intonation	Elements
1. Syntactic structure [a]	Speech paragraph
	Utterance
	Thought group
2. Mode	A: Assertion/statement/claim
	O: Order/directive
	Q: Question
3. Information structure	N/G: New/given
	Toward content (e.g., reflective, disclosing, ironic,)
4. Perceived attitude	Toward audience (e.g., earnest, ironic, convincing)

[a]. Vaissière proposed the syntactic structure function of intonation to be "the segmentation of continuous speech into syntactic units of different size" (p. 237). Accordingly, we identified three levels of analysis: the speech paragraph, utterance, and thought group

As shown in Table 1, the interpretative analysis of the second speech paragraph included, in the order indicated, syntactic structure (speech paragraph, utterance, thought group), mode (assertion, order, or question), information structure (new/given), and perceived attitude (toward content, toward audience).

The goal of the perceptual analysis was to identify the "local intonational phenomena" (Vaissière, 2005, p. 254). As shown in Table 2, these included the identification of key, pauses, and prominent and salient syllables.

Table 2

Perceptual Analysis Components

Intonational phenomenon	Elements
Key	High pitch to signifying new speech paragraph
Prominence (pitch, duration, intensity)	Syllable carrying focus/primary stress/nuclear accent
Salience (pitch, duration, intensity)	Somewhat prominent syllables or syllables carrying word-level stress
Tone choice	Falls, rises, level

Prominent syllables were identified as the most prominent syllable in a thought group, while salient syllables were perceived as less prominent. Chun (2002), quoting Cruttenden (1997), states that, "It is generally agreed that the three features of pitch, length, and loudness form a scale of importance in bringing syllables into prominence, with pitch being the most significant, duration next, and loudness the least important factor (cf. Cruttenden, p. 13)." Prominence was coded with bold capitalization (**FEW**) and salient syllables were underlined (e.g. learned) (see Table 3). Because of the importance of tone choice, "the prominent syllable on which the maximum, sustained pitch movement is identified" (Pickering, 2010) in discourse intonation (Cauldwell, 2015; Wells, 2006), this feature was also coded.

The acoustic analysis included measurement of pauses (Brown & Yule, 1983) and prominence (Boersma & Weenink, 2015; Chun, 2002; Cruttenden, 1997). According to Chun (2002), *pitch*, *length*, and *loudness* termed frequency, duration, and intensity, respectively are the physical properties of prominence. We measured prominence in terms of F0, duration and intensity of both prominent and salient syllables. Based on Figueroa (personal communication) and information obtained from the PRAAT manual, the acoustic measures were done using the dynamic menu, rather than editor window, to ensure greater reproducibility (Boersma & Weenink, 2015).

After completion of the separate analyses, a master spreadsheet was created to facilitate triangulated and cross-layer evaluation. There was considerable agreement amongst the researchers in most aspects of the interpretative and perceptual analyses. At the interpretative

level, at least two of the three researchers agreed on at least 96% of utterance and thought groups, mode, and informational structure judgments. Although the researchers used a range of different adjectives for the speaker's perceived attitude, there was always some agreement for both the attitude toward the content and toward the audience. At the perceptual level, at least two of the three researchers agreed on 100% of pause placement and prominence judgments. However, tone choice was eventually dropped from the analysis due to different interpretations and use of the coding scheme.

RESULTS

Speech paragraph 2 comprises seven utterances and 24 thought groups. These were analyzed to answer how Cutts' intonation contributed to the perceived impact of his TED Talk (RQ 1) and how these results inform the teaching of intonation (RQ 2).

Research Question 1: How does an interpretative, perceptual, and acoustic analysis of Matt Cutts' intonation elucidate the perceived impact of his TED Talk?

Interpretative and Perceptual Analysis

Table 3 presents the interpretative and perceptual analysis of all seven utterances. The perceptual analysis included key as well as prominent and salient syllables. Interpretative analysis includes thought groups (/), new/given information and perceived attitude of speaker toward the content and audience. Since all seven sentences were coded as assertions, mode is not listed.

Acoustic Analysis Results

The acoustic analysis added measurement of pauses, pitch range, prominent and salient syllables, as well as overall F0 and overall pitch contour. In the entire speech, mean pauses for utterances were 1.52 seconds while mean pauses for thought groups were .31 seconds demonstrating a general ratio of 1:5 (Table 4) between utterances and thought group pauses. These results show an expected role of pauses in parsing speech but more importantly a direct connection to the syntactic structure segmentation created through utterances and thought groups produced with consistent ratio within the parsing.

Table 3

Interpretative and Perceptual Analysis Results Summary for Speech Paragraph 2

#	Utterance/perceptual analysis	Interpretative analysis			
		New information	**Given information**	**Perceived attitude toward content**	**Perceived attitude toward audience**
1	//There's a **FEW** things I <u>learned</u> while doing these <u>thir</u>ty day <u>chall</u>enges//	few	while doing these thirty day challenges	reflective disclosing	sincere, relating
2	//The **FIRST** <u>was</u>/instead of the <u>months</u> <u>fly</u>ing by for**GOT**ten/the <u>time</u> was **MUCH**/**MORE**/**MEM**orable//	first, flying much, more, memorable		passionate	heartfelt
3	//**THIS** was part of a <u>chall</u>enge I did to take a <u>pic</u>ture/every <u>day</u> for a **MONTH**/and I remember e**XAc**tly/**WHERE** I <u>was</u>/and <u>what</u> I was **DO**ing that <u>day</u>//	take picture everyday exactly, where, what	part of a challenge for a month	factual	convincing
4	//I **AL**so <u>noticed</u>/that as I <u>started</u> to do **MORE** and **HAR**der thirty day <u>chall</u>enges/my self-**CON**fidence <u>grew</u>//	also noticed more and harder self confidence	thirty day challenges	proud	honest and heartfelt
5	//I went from **DESK** dwelling com<u>pu</u>ter nerd to/the <u>kind</u> of guy who **BIKES** to <u>work</u>/for **FUN**//	desk, computer kind of guy, bikes, fun		surprise, ironic	ironic
6	//**EV**en last <u>year</u>/I ended up hiking up <u>Mount</u> Kiliman**JA**ro/the **HIGH**est <u>moun</u>tain in <u>A</u>frica//	last year, Mount Kilimanjaro, highest		proud	inspiring
7	I would **NEV**er/have **BEEN**/that ad**VEN**turous/be**FORE** I <u>started</u> my <u>thir</u>ty day <u>chall</u>enges//	never, been adventurous	thirty day challenges	proud, disclosing	honest, inspiring, doable

Note. Results reported indicate agreement of at least two of the three researchers. Thought group boundaries (/); prominent syllables (**BOLD**); salient syllables (<u>underline</u>).

Table 4

Pause Results of Utterances and Thought Groups in the Entire Talk

	Sentence finals (seconds)	Thought groups (seconds)
Mean	1.52	0.31
Range	.78-1.90	.1-.80
Standard Deviation	0.41	0.20

In terms of pitch range, speech paragraph 2 was strongly representative of the entire talk (Table 5). The entire speech had a maximum pitch range of 489.1 Hz and minimum of 75.3 Hz with a 413.8 Hz spread. Consistency of pitch range seems indicative of its relevance for creating speech paragraphs and providing a clear structure for the audience to follow not only at the paragraph level but also across utterance and thought group levels, thus contributing to engaging discourse.

Table 5

Pitch Range Results by Level of Analysis

	Maximum (Hz)	Minimum (Hz)	Spread (Hz)
Entire Speech	489.13	75.3	413.8
Speech Paragraph (SP) 2	489.0	75.8	413.2
Utterances in SP2	491.1	78.5	412.6
Thought Groups in SP2	487.9	75.8	412.1

The results for pitch, duration, and intensity for prominent words and salient words (Table 6) indicated that F0 was the most differentiating indicator of these syllables. The mean pitch of prominent words was 200.44 Hz versus 144.46Hz for salient words, while duration and intensity varied little between the two categories.

Table 6

Prominent and Salient Syllable Measures

	Prominent syllables			Salient syllables		
	Pitch (Hz)	Duration (s)	Intensity (dB)	Pitch (Hz)	Duration (s)	Intensity (dB)
Mean	200.44	.12	69.24	144.56	.12	65.81
SD	46.43	.05	1.00	31.74	.04	4.49
Max	309.20	.24	71.40	236.20	.24	68.70
Min	140.50	.04	67.60	92.40	.05	49.10

Integrated Results for Utterance 1

Cutts starts speech paragraph 2 with a graphic organizer: "There's a few things I've learned while doing these 30 day challenges." Several things were immediately apparent from the interpretative analysis. The sentence is clearly an assertion/statement with FEW as new information and "while doing these thirty day challenges" as given information. The perceptual analysis of this utterance, shown in Table 3 above, ties to the interpretative structure through FEW being identified as prominent and the given information not being highlighted as prominently. The acoustic results revealed that F0 made the biggest difference between prominent and salient syllables with little variation in duration and intensity. As an example, the pattern for utterance 1 can be seen in Table 7.

Table 7

Measurements of Stress Variables in Prominent & Salient Words for Utterance 1

Words	Mean pitch (Hz)	Duration (s)	Intensity (dB)
FEW	278.00	0.04	70.80
learned	154.60	0.15	68.40
thir	137.60	0.10	67.80
chal	117.80	0.11	67.00

The results in Table 7 illustrate the marked difference in mean pitch between the word identified perceptually as most prominent (**FEW**) as compared to the words identified as salient in this utterance. In addition, the decreasing pitch over the course of the utterance is apparent in both the numbers and the PRAAT printout showing F0. Figure 1 shows the pitch contour with clear declination in this opening sentence, which was perceived as reflective and disclosing in Cutts' attitude toward the content as well as sincere and relating toward his audience.
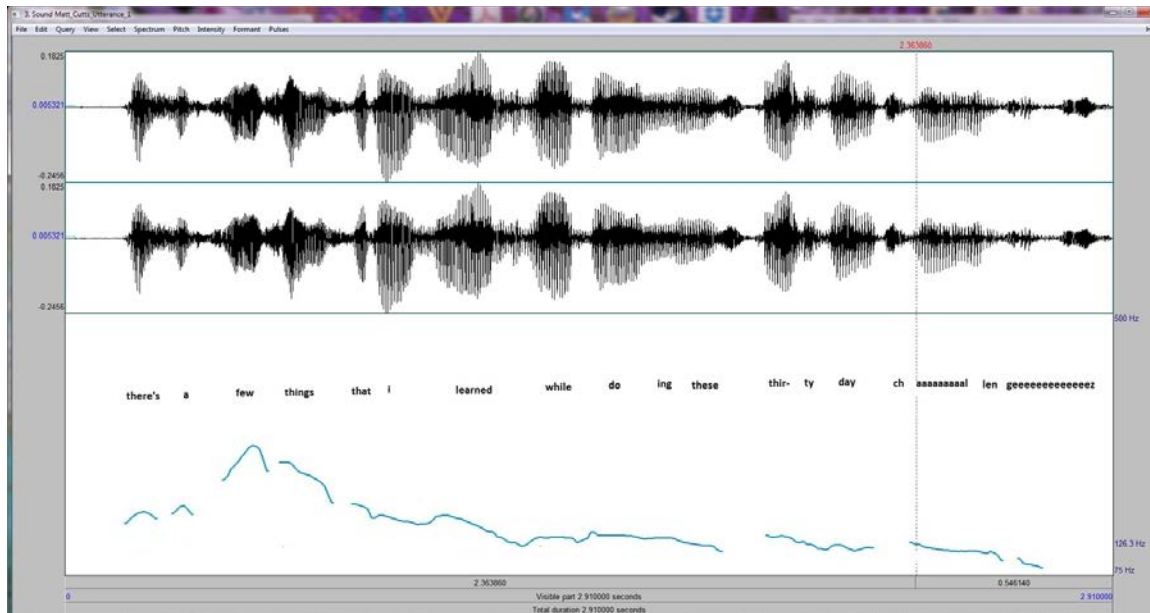
Figure 1. Acoustic analysis: speech paragraph 2-utterance 1

Research Question 2: How do the results of an interpretative, perceptual, and acoustic analysis of intonation inform teaching practice?

Three key pedagogical implications emerged. A sample syllabus based on these implications for teaching intonation using a TED Talk is included in Appendix B.

> 1) Start with meaning-making

TED Talks serve as authentic resources for more advanced level learners, and are examples of engaging monologic speech (Scotto di Carlo, 2014). Based on the layered approach to researching intonation of a TED Talk, we could apply this approach to enhance instructional effectiveness of using TED Talks for intonation training. Teachers could first focus students' attention on the engaging meaning-making; that is, the intended communicative outcome created by the speaker. Simply asking students what impression they get from a TED speaker begins the process of focusing students on the overall impression created. Following this with "how does the speaker do this so effectively?" forges language awareness (Borg, 1994) of both the discourse and the intonation features, and opens the door to the overlapping nature of these components. The paradigm shift for instructors might be from thinking of teaching intonation to thinking about teaching communicative effectiveness.

> 2) Use a layered approach

Instead of diving directly into the functions or parts of intonation, students can be guided in a type of noticing exercise (Schmidt, 1990) on overall organization, speech paragraphs, grammatical structures, formulaic language, and new versus given information, for example. Highlighting these aspects links meaning-making to the structures; this prepares students for connecting the next layer of intonation cues, which need to overlap simultaneously in oral production. After review of the language, teachers can scaffold working systematically through the role of intonation to parse and highlight (Sardegna & McGregor, 2013). Finally, students can use acoustic feedback from PRAAT to identify strengths and weaknesses in their executed intonation features or in comparison to a model.

> 3) Highlight the integrated systems

In addition to meaning-making and a layered approach, the arrows in Figure 2 show an interplay within the systems and also between their elements. Pauses create thought groups with prominent and salient syllables building pitch contours.
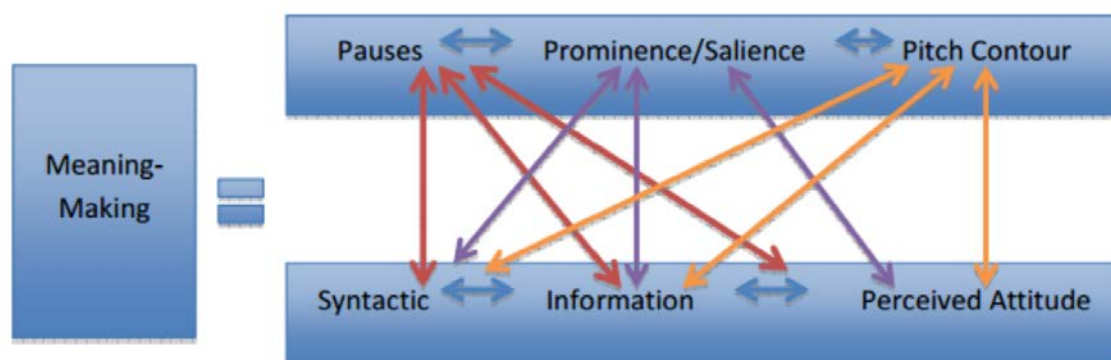
Figure 2. Integrated model of intonation

As shown in Table 4, utterance 1 represents a classic example in which the syntactic unit of a sentence overlaps the new/given information matching the peak of pitch contour on "few," the most prominent syllable. In contrast, we found atypical thought groups in utterance 2 in which pauses create one word thought groups with prominence for special emphasis.

> Utterance 2:   //The **FIRST** <u>was</u>/ instead of the <u>months</u> <u>fly</u>ing by for**GOT**ten/the <u>time</u> was **MUCH**/**MORE**/**MEM**orable//

Clearly, thought group production impacts prominence, which if inaccurately placed or produced will automatically change the overall pitch contour not to mention misalign with the syntactic and information structure and consequently impact the intended communicative effectiveness.

With the same foundational parts (thought groups, prominence, salience, tone choice, pitch range, pitch contours) of intonation, Cutts' perceived attitude changed from utterance 1 being reflective and sincere to utterance 2 showing his passionate and heartfelt attitude. The heartfelt, passionate, convincing attitude comes across by the dramatically short chunking, prominence on "for**GOT**ten" and contrast in the pitch, duration, and intensity at the end of the utterance. In order to achieve an advanced level of communication, students need skills to not only create syntactic units accurately, but also to use prominence accurately to produce thought groups that differentiate given from new information. Fluency development often poses a barrier with unintentional pauses not at syntactic units, which might cause breakdowns both at the language and intonation levels. Teachers need to understand the interplay and interactions between these systems to effectively provide instruction, scaffolded practice, and feedback on intonation.

DISCUSSION

Typically, when intonation is addressed in the classroom, it tends to become a description or list of parts (Meyers, 2014; Sardegna & McGregor, 2013) and teaching materials (see for example, Celce-Murcia et al., 2010), although rich in information, tend to adopt a parts and pieces view

of intonation instruction. The findings here support Levis's claim that intonation training should start with "…the primary goal of communicative proficiency rather than of teaching the mechanics of intonation…" (1999, p. 59). TED Talks are audience-oriented to appeal and engage (Scarlo di Carlo, 2014), and intonation was found in the present study to strongly contribute to how the speaker conveyed his attitude toward both the content and the audience. For these reasons, we recommend starting intonation instruction with meaning-making to go beyond intelligibility to comprehensibility where the learner can first recognize the outcome of communicative acts, be motivated to delve into the layers creating the meaning-making and finally focus on the interrelated features within the systems. An intonation toolbox of parts is insufficient and an integrated approach to learning about intonation embraces all aspects of the language strata (Halliday & Greaves, 2008) from the lexicogrammatical to phonological and phonetic.

This research was based on one sample of monologic speech by one native North American English speaker in a TED Talk forum. Although TED Talks are a rich and easily accessible online resource, certainly the data here are limited in generalizability. In addition to the aforementioned limitation, non-verbal communication was not systematically analyzed due to the limited camera angles of the video recorded material, but is recognized by the researchers as a critical layer for future research to explore. In spite of these limitations, the consideration of the 3-layer analysis and integrated approach to embrace the complexity of intonation is strongly encouraged for future research, teacher training, improved textbook development and classroom instruction. The goal, after all, is to equip students with more than just parts of intonation – it is to prepare them for their own capacity for communicative success and meaning-making.

ACKNOWLEDGMENTS

ABOUT THE AUTHORS

Alison McGregor holds a doctorate in Education Psychology and has been teaching English as Second Language for over 20 years. Currently, Alison is an instructor for the International Office at the University of Texas at Austin. She teaches advanced oral communication skills courses for international graduate students, ITAs and MBAs. In addition, she works as a communication consultant specializing in accent modification for the McCombs School of Business MBA program. Her research focuses on comprehensibility, factors that promote effective pronunciation instruction, and American English intonation. Alison can be reached at mcgregor@utexas.edu.

Beth Zielinski's research interests are in the area of L2 English speaking, particularly the influence of different pronunciation features on effective communication. Her research and teaching has involved learners of English in many different settings in Australia.

Colleen M. Meyers is an education specialist in the Center for Educational Innovation at the University of Minnesota. For over 30 years, she has taught in the International Teaching Assistant Program in addition to consulting with international faculty. A co-author of   Communicate: Strategies for International Teaching Assistants (2007) and English Communication for International Teaching Assistants (2012), in 2007 Colleen was awarded a Senior Fulbright scholarship to do teacher-training in the area of pronunciation teaching in Turkey.

Marnie Reed is Associate Professor of Education and affiliated faculty in the Program in Applied Linguistics at Boston University. She is the director of the graduate program in Teaching English to Speakers of Other Languages (TESOL) in the School of Education. Dr. Reed's research interests are applied phonetics and phonology, particularly the role of metacognition in cross-linguistic awareness of the pragmatic functions of intonation. She is co-author (with Christina Michaud) of Sound Concepts: An Integrated Pronunciation Course (McGraw-Hill), and co-editor (with John Levis) and chapter contributor to the Wiley-Blackwell Handbook of English Pronunciation.

REFERENCES

Baker, A., & Murphy, J. (2011). Knowledge Base of Pronunciation Teaching: Staking Out the Territory. *TESL CANADA JOURNAL, 28*(2).

Boersma,P., & Weenink, D. (2015). Praat: Doing phonetics by computer. Retrieved September 2015 from http://www.praat.org/.

Borg, S. (1994). Language awareness as a methodology: Implications for teachers and teacher training. *Language Awareness 3*(2), 61–71.

Brown, G. & Yule, G. (1983). *Discourse Analysis*. Cambridge Textbooks in Linguistics. Cambridge University Press.

Cauldwell, R. (2015). *Discourse intonation*. Retrieved from: http://www.speechinaction.org/for-teachers/discourse-intonation/

Celce-Murcia, M., Brinton, D., & Goodwin, J. (2010). *Teaching Pronunciation: A Course Book and Reference Guide* (2nd ed.). New York, NY: Cambridge University Press.

Chase, B T., Milner, M., & Johannsen, K. L. (2015). *World English 1* (2nd ed.). Boston: National Geographic, Cengage Learning.

Chun, D. (2002). *Discourse Intonation in L2: From Theory & Research to Practice*. Benjamins, John Publishing Company.

Cruttenden, A. (1997). *Intonation.* 2nd edition. New York: Cambridge University Press.

Cutts, M. (2011). *Try something new for 30 days* [Video file]. Retrieved from: http://www.ted.com/talks/matt_cutts_try_something_new_for_30_days?language=en

Halliday, M. A. K., & Greaves, W. S. (2008). *Intonation in the Grammar of English*. London: Equinex Textbooks and Surveys in Linguistics.

Hincks, R. (2005). Measures and perceptions of liveliness in student oral presentation speech: A proposal for an automatic feedback mechanism. *System, 33*(4), 575-591.

Hincks, R. & Edlund, J. (2009). Promoting increased pitch variation in oral presentations with transient visual feedback. *Language Learning & Technology,* 13(3), 32-50.

Levis, J. (2005). Comparing Apples and Oranges? Pedagogical Approaches to Intonation in British and American English In Katarzyna Dziubalska-Kotaczyk & Joanna Przedlacka (eds), English Pronunciation Models: A Changing Scene (pp. 339 - 366) Bern: Peter Lang.

Levis, J. (1999). Intonation in Theory and Practice, Revisited. *TESOL Quarterly, 33*(1), 37-63

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

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

Meyers, C. (2014). Intelligible accented speakers as pronunciation models. In J. Levis & S. McCrocklin (Eds). Proceedings of the 5th Pronunciation in Second Language Learning and Teaching Conference (pp. 172-76). Ames, IA: Iowa State University.

Paunovic, T., & Savic, M. (2008). Discourse intonation - Making it work. English Language Overseas Perspectives and Enquiries, 5(1-2), 57-75.

Pickering, L. (2001). The role of tone choice in improving ITA communication in the classroom. *TESOL Quarterly 35*(2), 233-255.

Pickering, L. (2004). The structure and function of intonational paragraphs in native and non-native speaker instructional discourse. *English for Specific Purposes, 23*(1), 19-43.

Pickering, L. (2010). Intonation. In K. Malkjaer (Ed.), *The Linguistics Encyclopedia* (3rd ed., pp. 283-289): Routledge.

Reed, M. & Michaud, C. (2015). Intonation in research and practice: The importance of metacognition. In M. Reed & J. M. Levis (Eds.), *The Handbook of English Pronunciation* (pp. 454-470). West Sussex, UK: John Wiley & Sons, Inc.

Sardegna, V., & McGregor, A. (2013). *Scaffolding students' self-regulated efforts for effective pronunciation practice.* Paper presented at the Pronunciation in Second Language Learning and Teaching Conference, Ames, IA: Iowa State University.

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2),129–58.

Scotto di Carlo, G. (2014). The role of proximity in online popularizations: the case of TED talks. *Discourse Studies, 16*(5), 591-606.

Slater, T., Levis, J., & Levis Muller, G. (2015). Spoken parentheticals in instructional discourse in STEM and Non-STEM disciplines: The interaction of the prosodic, ideational, and interpersonal resources in signaling information structure. In G. Gorsuch (Ed.), *Talking matters: Research on talk and communication of international teaching assistants*. Stillwater, OK: New Forums Press.

Vaissière, J. (2005). Perception of intonation. In D.B. Pisoni & R.E. Remez (Eds.), *The Handbook of Speech Perception* (pp. 236-263). Malden, MA: Blackwell Publishing.

Wells, J.C. (2006) *English Intonation: An Introduction*, Cambridge University Press.

Appendix A

Transcript with speech paragraphs (///), utterances (//) and thought groups (/) marked

Speech paragraph 1

a few years ago/ I felt like I was stuck/ in a rut// so I decided to follow in the footsteps of the great/American/ philosopher/ Morgan Spurlock/ and try something new/ for thirty days// the idea is actually pretty simple// think about something you've always wanted to add to your life/ and try it/ for the next thirty days// it turns out/ thirty days is just about the right amount of time/ to add a new habit/ or / subtract a habit/ like watching the news/ from your life///

Speech paragraph 2

there's a few things I learned while doing these thirty day challenges// the first was/ instead of the months flying by forgotten/ the time was much/ more/ memorable// this was part of a challenge I did to take a picture/ every day for a month/ and I remember exactly/ where I was/ and what I was doing that day// I also noticed/ that as I started to do more and harder thirty day challenges/ my self confidence grew// I went from desk dwelling computer nerd to/ the kind of guy who bikes to work/ for fun// even last year/ I ended up hiking up Mount Kilimanjaro/ the highest mountain in Africa// I would never/ have been / that adventurous / before I started my thirty day challenges///

Speech paragraph 3

I also figured out/ that/ if you really want something badly enough/ you can do anything/ for thirty days// have you ever wanted to write a novel// every November/ tens of thousands of people/ try to write their own fifty thousand word novel /from scratch/ in thirty days// it turns out/ all you have to do/ is write sixteen hundred and sixty-seven words a day/ for a month// so I did// by the way the secret/ is not to go to sleep/ until you've written your words for the day// you might be sleep deprived/ but/ you'll finish your novel// now/ is my book the next great American novel// no/ I wrote it in a month// it's awful// but/ for the rest of my life/ if I meet John Hodgman at a TED party/ I don't have to say/ I'm a computer scientist// no no/ if I want to I can say/ I'm a novelist///

Speech paragraph 4

so here's one last thing I'd like to mention// I learned that when I made small/ sustainable changes/ things I could keep doing/ they were more likely to stick// there's nothing wrong with big crazy challenges / in fact/ they're a ton of fun/ but they're less likely to stick// when I gave up sugar/ for thirty days/ day thirty-one looked like this///

Speech paragraph 5

so here's my question to you //what are you waiting for// I guarantee you the next thirty days/ are going to pass/ whether you like it or not// so why not think about something you have always / wanted / to try/ and give it a shot/ for the next thirty days//

Appendix B

Four-week Mirroring Project Syllabus

Week	Analysis	Intonation focus	In-class Activity	Homework
1	Context-setting & interpretative analysis	Thought groups & boundary tones	Discussion of topic Connection to what undergraduates expect Mark transcript for thought groups & pausing	Listen, view & mark focus words & body language
2	Perceptual analysis	Prominent and salient words Body language	Compare marking in pairs. View video to compare marking. Work with PRAAT recordings.	"Mirror" video
3	Acoustic analysis (PRAAT)	Tone choice	Work with PRAAT recordings. Record "cold" version	Critique "cold" version
4	Interpretation	ALL	Record "final" version	Complete self-critique form

Richards, M. (2016). Transforming any text into an individualized segmental exercise via RelateWorldwide's *Pronunciation Highlighter*. 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. 160-167). Ames, IA: Iowa State University.

TRANSFORMING ANY TEXT INTO AN INDIVIDUALIZED SEGMENTAL EXERCISE VIA RELATEWORLDWIDE'S *PRONUNCIATION HIGHLIGHTER*

Monica Richards, Iowa State University

In many L2 English pronunciation classes, it is difficult for teachers to provide segmental instruction and practice adequate for enabling students to acquire consistently accurate pronunciation of all high-functional-load segmentals they find challenging (Munro & Derwing, 2006). In part, this is because students in ESL classes rarely all need instruction and practice for the same English segmentals (Swan & Smith, 2001). While freely available asynchronous segmental training online may seem a natural solution to this problem and while outstanding segmental instruction has been available online for some time, few substantive practice resources have been available until recently to help students make their pronunciation fluent and automatic. This paper introduces the author's new online Pronunciation Highlighter, a tool capable of transforming any text students find interesting (textbook dialogues, technical term lists, PowerPoint presentation outlines, etc.) into individualized segmental practice exercises. The paper also suggests ways students can capitalize on Pronunciation Highlighter output to build new and accurate segmental pronunciation habits.

INTRODUCTION

Segmentals (the individual consonant and vowel sounds of a language), and particularly high-functional-load[1] segmentals (Brown, 1988, 1991; Catford, 1987; King, 1967; Munro & Derwing, 2006), can make a difference in the intelligibility and comprehensibility[2] of spoken English to NNS and NS listeners (Jenkins, 2000; Munro & Derwing, 2006). Unfortunately, as Jenkins (2000) points out, "It is widely agreed that habit formation in language transfer figures more extensively at the

---

[1] Functional load is defined by King (1967) in relation to phonology as "a measure of the work which two phonemes (or a distinctive feature) do in keeping utterances apart" (p. 831). Brown (1988) indicates that two important means of measuring functional load are (1) the frequency of minimal pairs containing any two given phonemes and (2) the frequency of each phoneme individually in the target language. Munro and Derwing (2006) suggest an additional relevant factor may be a target phoneme's position in a particular word since research indicates that "word-initial errors are more important than errors in other positions" (p. 530).

[2] The terms "intelligibility" and "comprehensibility" are here used in their technical sense, following Munro and Derwing (2006), with the term "intelligibility" referring to the degree to which a listener *successfully* understands what a speaker says and the term "comprehensibility" referring to listeners' impression of how *easy* it is to understand what a speaker says.

phonological level than at either the syntactic or lexical levels. . . . The production of speech sounds is unlike that of lexis and syntax, since it does not involve passing messages through the brain, but rather the development of highly automatized motor skills and, consequently, over time, the formation of L1 speech habits which are not easily de-automatized in L2" (Jenkins, 2000, p. 112). That is, while English language learners may gain explicit knowledge of how English segmentals are articulated through interesting and memorable pronunciation pedagogy such as that instantiated in Marsha Chan's *Pronunciation Doctor* YouTube videos: (https://www.youtube.com/channel/UCmo0sgPqUCPDLcGhi-J_JEg), facilitating the interface of this explicit knowledge into students' implicit, everyday habits of speaking English remains challenging. This is particularly true for cases such as that in which two phones (e.g., /n/ and /l/) are allophonic in a learner's L1 (e.g., the Sichuan/Hunan dialects of Chinese), but in English are full phonemes that therefore appear inconsistently "sprinkled" across the learner's L2 English speech, producing — where the problem phones carry high functional load as /n/ and /l/ do in English — notoriously low-comprehensibility or even unintelligible L2 English speech that often proves difficult to modify (Jenkins, 2000; Richards, 2012).

Segmentals matter for L2 English learners' English *listening*

Jenkins (2000) suggests that in L1-allophone/L2-phoneme cases such as for /n/ and /l/ above "Although production would . . .prove difficult prior to extensive instruction, it would almost certainly precede perception. We cannot assume that because [nonbilingual English speakers] are able to produce sound contrasts, they can necessarily discriminate aurally between them; indeed, the opposite is not infrequently true" (Jenkins, 2000, pp. 33-34). Jenkins' statement reflects Ladefoged's (1967) claim that "acoustic differences cannot be readily perceived until the corresponding articulatory gestures have been learnt" (p. 167). That is, unless students learn first to *articulate* the difference between English segmentals, they are unlikely to acquire the *perceptual* ability to distinguish them.

Failure to distinguish English segmentals aurally not only results in learners' inability to differentiate pseudohomophones they hear (e.g., "glass" and "grass" for L1 Japanese listeners), but far more seriously, results in their continuing to face spurious activation of pseudoembedded words as they attempt processing the stream of speech, with the result that their word segmentation and word identification processes remain inefficient. Cutler (2012) dramatically illustrates the importance of this concern by examining the high-functional-load /r/ vs. /l/ distinction, indicating that "frequency-adjusted estimates [based on the CELEX English Lexicon]. . . suggest 49,508 spurious embeddings per million words due to misperceiving [l] as [r], and 69,923 per million words for misperception of [r] as [l]" (p. 322).

The degree to which learners are helped in combating these segmentally-grounded L2 English listening problems by increasing their *explicit* knowledge of how problem English segmentals are articulated *when this explicit knowledge has not yet been proceduralized* remains an open question for further research. In all probability, however, learning to *habitually* articulate in near-standard form at least high-functional-load segmentals is important not only for increasing L2 English learners' spoken English intelligibility and comprehensibility but also for improving their English listening ability.

Can current resources proceduralize near-standard English segmental pronunciations?

Ellis (2008) provides helpful definitions of the implicit and explicit knowledge constructs in cognitive linguistics, defining *implicit knowledge* as being "intuitive, procedural, systematically variable, automatic, and thus available for use in fluent, unplanned language use." In contrast, Ellis defines *explicit knowledge* as being "conscious, declarative, anomalous, and inconsistent (i.e. [explicit knowledge] takes the form of 'fuzzy' rules inconsistently applied) and [is] generally only accessible through controlled processing in planned language use" (p. 418). While there is disagreement regarding whether or not implicit linguistic knowledge can be acquired post-puberty (Ellis, 2008) and whether the implicit/explicit knowledge distinction represents a continuum or dichotomy (Dienes and Perner, 1999), certainly it *is* important that learners be helped to proceduralize/automatize as much as possible near-standard articulation of high-functional-load segmental contrasts.

Many researchers (and, by definition, arguably all pronunciation *teachers*!) subscribe to a model of instructed L2 acquisition assuming the potential of at least weak interface between learners' explicit and implicit knowledge that allows the conversion of explicit knowledge to implicit knowledge at some point and to some degree (Ellis, 2008) — usually by means of practice. Unfortunately, a cursory review of pronunciation textbooks and software suggests that current offerings fail to provide learners with the quantity of practice likely necessary to assist the proceduralization process. It is to fill this gap that I have developed my Web-based *Pronunciation Highlighter*. This tool not only enables materials development for providing learners the varying amounts of practice likely to help proceduralize their standard/near-standard articulation of any given English segmental, but also allows learners freedom to choose the *content* that they input, with the goal of maximizing their interest in practicing their specific problem segmentals.

Overviewing RelateWorldwide's *Pronunciation Highlighter*

The *Pronunciation Highlighter* introduced above is available on the author's RelateWorldwide (**Re**sources for the **l**earning **a**nd **t**eaching of **E**nglish worldwide) website at http://www.relateworldwide.org/speaking/pronunciation-highlighter/. The *Pronunciation Highlighter* allows a student or teacher to input any text (Figure 1), indicate problem phonemes the *Highlighter* should identify and adjust the tool's default highlighter colors as desired. The design of the *Pronunciation Highlighter* is informed by pronunciation research findings in that its phoneme-selection component informs users of phonemes' high- and low-functional-load status (Brown, 1988, 1991; Catford, 1987; King, 1967; Munro & Derwing, 2006) by marking them either (+) or (-) respectively, thereby encouraging users to prioritize phonemes carrying high functional load and de-emphasize those carrying low functional load. When the user clicks "Submit," the *Highlighter* outputs his or her text with all words containing problem phonemes (according to version 0.2 of the freely available, redistributable *Illinois Speech and Language Engineering Dictionary* containing

Figure 1. RelateWorldwide's *Pronunciation Highlighter* input box for user-chosen input texts

137,000 words — Hasegawa-Johnson, 2007[3]) highlighted in the colors chosen. Words containing multiple problem phonemes are highlighted in yet another color — gray by default — to inform the user that these words are likely to need particularly careful attention and practice. Below the highlighted version of the user's input text, the *Pronunciation Highlighter* displays a list of all words in the text containing the user's problem phonemes (Figure 2). Each listed word is

---

[3] Because RelateWorldwide's *Pronunciation Highlighter* currently utilizes only the *Illinois Speech and Language Engineering Dictionary* (which is 90% based on [Hasegawa-Johnson, 2007] the *Carnegie Mellon University Pronouncing Dictionary,* "an open-source machine-readable pronunciation dictionary for North American English" — Lenzo, n.d.), its pronunciation highlighting currently reflects only North American English pronunciation norms.

hyperlinked to its YOUGLISH YouTube video "playlist," where each video in the "playlist" is cued to just a few seconds before the target word is spoken.[4]

Learning and teaching with RelateWorldwide's *Pronunciation Highlighter*

While RelateWorldwide's *Pronunciation Highlighter* has potentially several research and teaching applications, I highlight a few possibilities below to hopefully serve as fodder for creative pronunciation researchers, English language teachers and English pronunciation students to develop additional innovative ways of applying the *Pronunciation Highlighter* to maximize the ease with which English language learners can practice — with the aim of proceduralizing — standard/near-standard articulation of particularly high-functional-load English segmentals.

Inputting textbook conversation transcripts. Perhaps the most obvious application of the *Pronunciation Highlighter* is for teachers and students to input into the *Highlighter* the ubiquitous dialogues found in English as an international language (EIL) and ESL listening, speaking and pronunciation textbooks[5] and then to select first for practice challenging high-functional-load, and later, if desired, low-functional-load segmentals (Brown, 1988, 1991; Catford, 1987; King, 1967; Munro & Derwing, 2006).[6] Due to limits in language learners' attention/working memory resources (Ellis, 2008), it is probably advisable in most cases that 1) fewer than the *Pronunciation Highlighter*'s maximum of five problem segmentals be selected and 2) students first work on articulating their problem segmentals in the context of each segmental's output word list and only after that in the context of the *Pronunciation Highlighter*'s highlighted version of their original input text.

---

[4] YOUGLISH is a resource for unscripted, contextualized stream-of-speech intensive English listening/pronunciation practice hyperlinking 300,000+ English word and name spellings to a playlist of their appearances in what are apparently — based on closed-captioning quality — human-transcribed YouTube videos: http://youglish.com/. In addition to its default "all [dialects]" display option, YOUGLISH allows search results to be limited to the — almost certainly loosely defined — "US" and "UK" dialect families. YOUGLISH also allows phrase searches.)

[5] Ideally, most dialogues, etc., input into the *Pronunciation Highlighter* should instantiate formulaic speech routines characteristic of informal (everyday) or formal (academic and professional) English communication. If , as sometimes happens, textbook examples fail to instantiate functional language patterns and phraseology characteristic of standard English as it is really spoken in EIL/ESL contexts (Folse, 2004), students may be at risk of automatizing comprehensibility-detracting nonstandard phrasing and grammar. Example supplemental resources that *do* generally instantiate standard formulaic speech routines, phrasing and grammar (Derewianka, 1990, 2011; Folse, 2004; Francis, Hunston & Manning, 1996, 1998; O'Keeffe, McCarthy & Carter, 2007) include http://openlanguage.com/library/learn-english/10/complimentary-english-course/13 (for everyday communication) and https://www.ted.com/talks (for academic lecture/professional presentation communication).

[6] Particular learners' problem segmentals can best be identified through a diagnostic assessment of their ability both to *perceive* the various English segmentals (e.g., via a minimal-pair discrimination task) as well as to *produce* them (e.g., via reading aloud a standardized diagnostic passage containing most or all the segmentals/consonant clusters of English in the various syllable positions each can appear along with production of a free speech sample on some familiar topic) (Celce-Murcia, Brinton & Goodwin, 1996). Where individual diagnostic evaluations such as the above are not possible, a reference detailing the common English segmental errors characteristic of speakers from various L1s (e.g., Avery & Ehrlich, 1992; Kenworthy, 1987; Swan & Smith, 2001) can be used in conjunction with information on functional load (Brown, 1988, 1991; Catford, 1987; King, 1967; Munro & Derwing, 2006) to identify segmentals likely worthy of focus.

Inputting technical term lists. Mispronunciation of technical terms associated with one's academic discipline or profession can not only result in reduced intelligibility and comprehensibility, but also negatively impact listeners' impression of one's academic and professional competence. Proceduralizing near-standard pronunciation of the technical terms in one's field, therefore, should be a top priority of all learners of L2 English for academic or professional purposes.

Inputting formal presentations' "slide" text. In many contexts, we cannot predict what words we will need to say. However, in the often high-stakes context of formal presentations, one's PowerPoint or other presentation "slides" frequently contain most of the key vocabulary needed to present one's topic. Practicing standard/near-standard pronunciation of each slide's text, therefore, followed by repeated, semi-spontaneous practice through one's entire "slide" presentation is likely to maximize the intelligibility and comprehensibility of L2 English speakers' presentations.
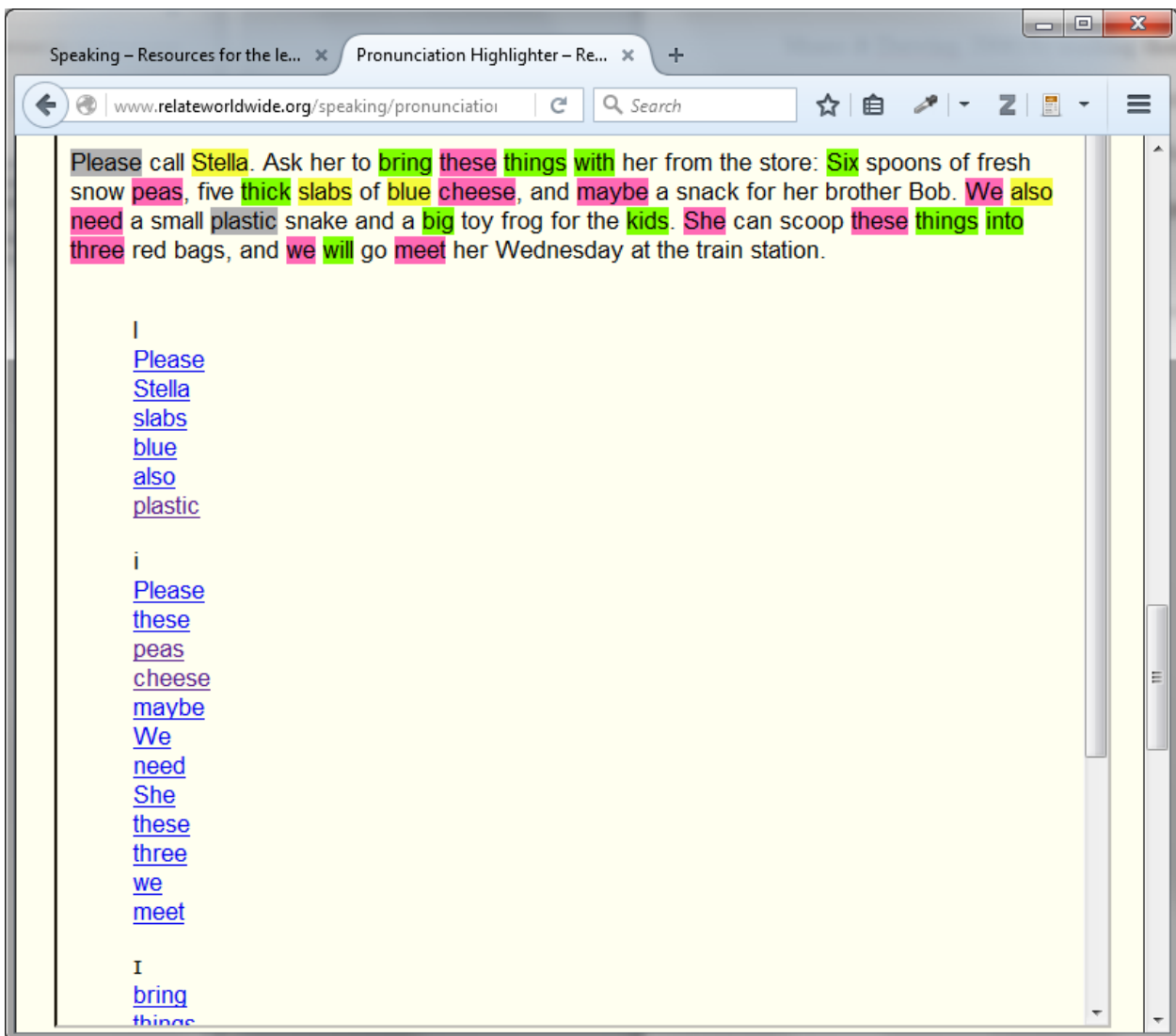
Figure 2. Example RelateWorldwide *Pronunciation Highlighter* output for a user-chosen input text

CONCLUSION

RelateWorldwide's *Pronunciation Highlighter* has been designed to enable L2 English learners and teachers to create high-interest materials in sufficient quantity that learners can engage in the practice required to proceduralize standard/near-standard articulation of English segmentals, particularly high-functional-load segmentals. RelateWorldwide's *Pronunciation Highlighter* can thus help L2 English learners reach their intelligibility and comprehensibility goals.

ABOUT THE AUTHOR

Monica Richards is a Ph.D. candidate in Applied Linguistics and Technology at Iowa State University. Her research interests focus on the intersection of psycholinguistics and English language listening and pronunciation, English-as-an-international-language pedagogy, and CALL (computer-assisted language-learning), especially among Chinese learners of English. **Contact information:** 443 Ross Hall, Iowa State University, Ames, IA 50011. 515-450-4169, monicagr@iastate.edu

REFERENCES

Avery, P., & Ehrlich, S. (1992). *Teaching American English pronunciation.* Oxford, UK: Oxford University Press.

Brown, A. (1988). Functional load and the teaching of pronunciation. *TESOL Quarterly, 22*, 593-606.

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

Catford, J.C. (1987). Phonetics and the teaching of pronunciation: A systemic description of English phonology. In J. Morley (Ed.), *Current perspectives on pronunciation: Practice anchored in theory* (pp. 86-99). Washington, DC: TESOL.

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

Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words*. Cambridge, MA: The MIT Press.

Derewianka, B. (1990). *Exploring how texts work.* Newtown, NSW, Australia: Primary English Teaching Association.

Derewianka, B. (2011). *A new grammar companion for teachers (2nd ed.).* Marrickville Metro, NSW: Primary English Teaching Association Australia.

Dienes, Z., & Perner, J. (1999). A theory of implicit and explicit knowledge. *Behavioral and Brain Sciences*, *22*(5), 735–808.

Ellis, R. (2008). *The study of second language acquisition (2nd ed.).* Oxford, UK: Oxford University Press.

Folse, K. (2004). *Vocabulary myths: Applying second language research to classroom teaching.* Ann Arbor, MI: The University of Michigan Press.

Francis, G., Hunston, S. & Manning, E. (Eds.) (1996). *Collins COBUILD grammar patterns 1: Verbs — Helping learners with* real *English.* London, UK: HarperCollins Publishers.

Francis, G., Hunston, S. & Manning, E. (Eds.) (1998). *Collins COBUILD grammar patterns 2: Nouns and adjectives — Helping learners with* real *English.* London, UK: HarperCollins Publishers.

Hasegawa-Johnson, M. (2007). Illinois Speech and Language Engineering Dictionary (ISLEdict) (Version 0.2) [Software]. Available from https://github.com/jart/poemy2/tree/master/data/isledict

Jenkins, J. (2000). *The phonology of English as an International Language: New models, new norms, new goals.* Oxford, UK: Oxford University Press.

Kenworthy, J. (1987). *Teaching English pronunciation*. New York, NY: Longman.

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

Ladefoged, P. (1967). Three areas of experimental phonetics: Stress and respiratory activity, the nature of vowel quality, units in the perception and production of speech. Oxford, UK: Oxford University Press.

Lenzo, K. (n.d.) Carnegie Mellon University Pronouncing Dictionary (CMUdict) [Software]. Available from http://www.speech.cs.cmu.edu/cgi-bin/cmudict

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

O'Keeffe, A., McCarthy, M. & Carter R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge, UK: Cambridge University Press.

Richards, M. (2016). *Pronunciation highlighter (based on North American English).* Retrieved from http://www.relateworldwide.org/speaking/pronunciation-highlighter/

Richards, M. (2012). Helping Chinese learners distinguish English /l/ and /n/. In. J. Levis & K. LeVelle (Eds.). Proceedings of the 3rd Pronunciation in Second Language Learning and Teaching Conference, Sept. 2011. (pp. 161-167). Ames, IA: Iowa State University.

Swan, M., & Smith, B. (2001). *Learner English: A teacher's guide to interference and other problems*. Cambridge, UK: Cambridge University Press.

Willis, D. (2003). *Rules, patterns and words: Grammar and lexis in English language teaching.* Cambridge, UK: Cambridge University Press.

Talley, J. (2016). What makes a Bostonian sound Bostonian and a Texan sound Texan? 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. 168-179). Ames, IA: Iowa State University.

WHAT MAKES A BOSTONIAN SOUND BOSTONIAN AND A TEXAN SOUND TEXAN?

Jim Talley, Linguistic Computing Systems, Austin, TX

This paper introduces a preliminary version of a new methodology for the automated, data-driven discovery of acoustic features of speech which potentially contribute to an accent's distinctiveness. The results discussed herein, while merely illustrative at this stage, provide reason to be optimistic about the prospects of evolving a truly useful and robust automated methodology for cataloging the characteristic acoustic aspects of accented speech. If this line of research were to fully fulfill its promise, the resulting comprehensive catalog of features would contribute to our explicit knowledge of the correlates of accent. The knowledge represented by such a catalog could potentially be directly applied by teachers of second language pronunciation, and it certainly would inform the development of the more capable and individualized computer-assisted pronunciation training (CAPT) tools of the future.

INTRODUCTION

At some level, all of us are aware of accents in speech in our native languages. We distinguish accents (acoustic characteristics due to the provenance or linguistic background of a speaker) from other acoustic idiosyncrasies of a speaker (such as those due to an individual's physical characteristics). With varying degrees of skill, some of us are able to identify the first language (L1) of a non-native speaker or the dialectal region of a native speaker (at least some of the time). This is a fairly difficult task, and harder yet is the task of identifying (especially in real-time) the acoustic features of speech which cause it to be perceived as accented. Yet this challenging task is an implicit requirement for teachers of second language (L2) pronunciation, since knowing what is making students' speech sound accented is a prerequisite for explaining to them how to sound less accented. The task is challenging enough when a teacher is intimately familiar with the L1(s) of his/her students, and still more challenging when the L1s of students are unfamiliar.

A comprehensive catalog of L2 pronunciation issues commonly exhibited by speakers of a specific L1 can be a useful tool for L2 pronunciation teachers, as long as teachers are cognizant of the fact that it merely provides an enumeration of *possible* pronunciation issues to watch for, rather than predicting exactly the pronunciation errors that all speakers with that L1 background will make. Neri, Cucchiarini, Strik & Boves (2002) also highlights the need for such knowledge in computer-assisted pronunciation training (CAPT). Discussing one of the most advanced

168

CAPT systems (the ISLE project), they say "this approach can only be adopted for specific L1-L2 pairs for which sufficient knowledge of typical pronunciation errors is available" (p.457).

Swan & Smith (2001) represents perhaps the most comprehensive attempt to manually catalog likely L2 pronunciation (and other) issues for a significant range of L1 languages. Derwing & Munro (2015, p.72) takes issue with Swan & Smith's "global prediction of difficulty," but this criticism is less well founded if (as discussed above) such a catalog is viewed as an overly large set of *possible* L1-sourced issues. Derwing & Munro rightfully emphasize the individual variability with respect to actual pronunciation issues. Individualization is where a CAPT system, based on an extensive catalog of potential issues, could be well-equipped to shine. The (very preliminary) research reported on in this paper takes an approach quite unlike the human expert based cataloging of Swan & Smith. This complementary approach is a bottom-up method, starting from machine-detectable, acoustic features derived from corpora of recorded speech. It uses a newly elaborated, machine learning (ML) based methodology to attempt to automatically create a catalog of the characteristics which distinguish one speaker population from another (*e.g.*, Quechua speakers who are learning English vs. native English speakers). Since this knowledge is automatically derived directly from base recordings, it follows that it would be detectable (and actionable) in a CAPT framework.

The Speech Data

The method discussed below is generally applicable to characterizing accent differences between sub-populations of speakers given a representative corpus of speech upon which to train. The work described herein focuses on learning the distinguishing characteristics of regional dialects of American English, rather than, say, distinguishing characteristics for Malayalam-speaking learners of English, for no better reason than that the necessary type of training data was readily available in the form of the TIMIT database[1].

The TIMIT speech database (Garofolo, et al., 1993) consists of clean (laboratory) recordings from 630 speakers (70% male, 30% female, of varied ages). The speakers were categorized into 7 dialect regions (DRs) – New England, Northern, North Midland, South Midland, Southern, NYC, and Western – based on where they had grown up. It also defines an "Army Brat" pseudo-region for those who lived in multiple DRs during childhood, presumably becoming speakers of Standard American English (SAE). Figure 1, based upon a photo included in Garofolo, et al. (1993), illustrates the 7 geographical TIMIT dialect regions.

We do not necessarily endorse the choice of these DRs as being ideal, neither do we assume homogeneity within each DR – they are simply all that we have to work with. It is some consolation, however, that the TIMIT DRs correspond fairly well with the 6 DRs delineated by Labov, Ash, & Boberg (2006). We should also note that the TIMIT speakers were not selected for, nor evaluated on, the prototypicality of their regional accents, and many may have effectively been speakers of SAE rather than true DR representatives.

---

1   TIMIT is distributed by the Linguistic Data Consortium (www.ldc.upenn.edu).
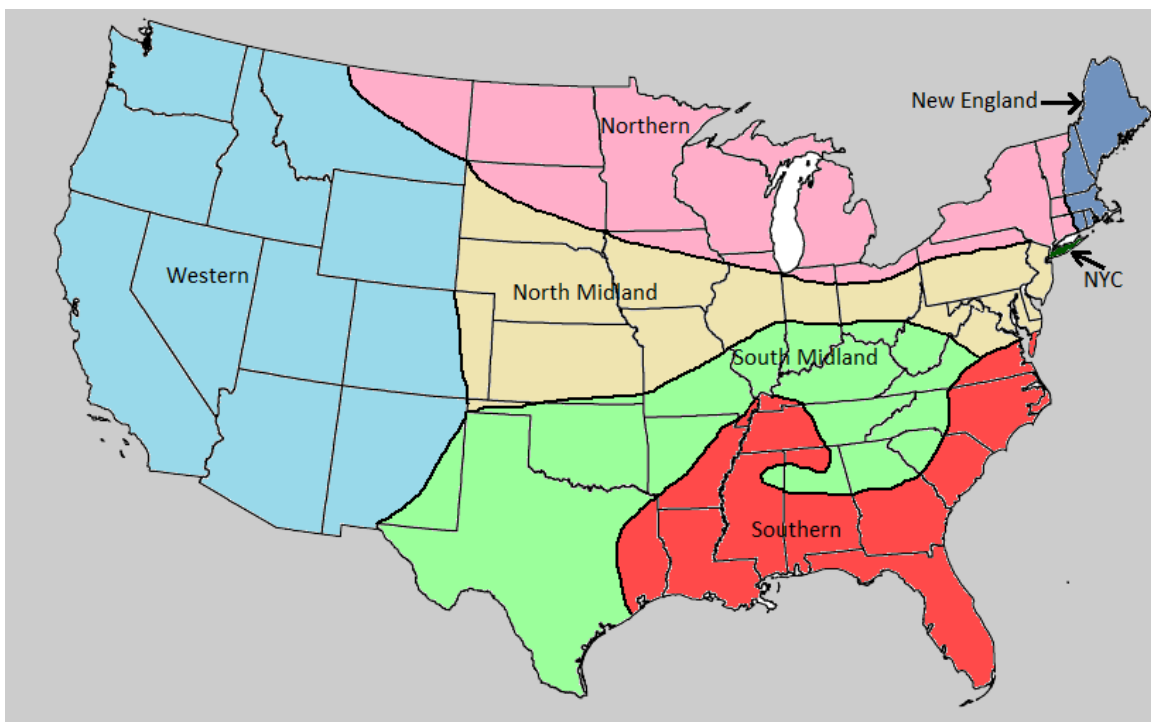
Figure 1. The TIMIT dialect regions (DRs).

There are three classes of recorded sentences in the TIMIT database – Dialect (SA), Compact (SX), and Diverse (SI). The two SA sentences (see Figure 2) were designed to expose differences in regional dialects and were recorded by each of the 630 TIMIT speakers. There were 450 distinct SX sentences, designed for phonetic diversity and coverage. Each of the speakers recorded 5 SX sentences, and each SX sentence was recorded by 7 speakers. The SI sentences were selected for phonetic context diversity from existing sources. Each TIMIT speaker recorded 3 unique SI sentences. For our purposes, we divided the corpus sentences into two groups: the 8 SI & SX sentences from each speaker vs. the 2 SA sentences from each speaker. We used the SI/SX subset for training of the front-end (FE) system, and the SA subset was used for back-end (BE) training and testing.

Figure 2. The (canonically transcribed) SA sentences from the TIMIT database.

Research Goals

The desired end goal for this line of research is to comprehensively catalog features of speech which distinguish accents of sub-populations. Such a catalog would serve to inform the development of CAPT tools and could serve as an additional resource for pronunciation training

practitioners. This paper provides an outline of a still underdeveloped methodology for achieving those goals and, as such, it has the modest objectives of 1) providing indications that it has the potential to be developed into an effective technique for meeting the ultimate goal, and 2) exposing the methodology to others in the field for vetting, feedback, and elaboration.

METHODS

This section describes how, starting from raw speech data plus transcriptions, we arrive at ranked lists of features for dialect identification. It has two primary blocks of ML (dubbed the Front-End [FE] and Back-End [BE]) which are connected by a number of data transformation steps (the "Glue").

Front-End Machine Learning

The objective of the FE ML is to learn the models (neural network [NN] and hidden Markov model [HMM]) which enable us to convert from digital speech recordings into temporally segmented frames of descriptive features, with segmentation conforming to phone[2] boundaries (see Figure 3).
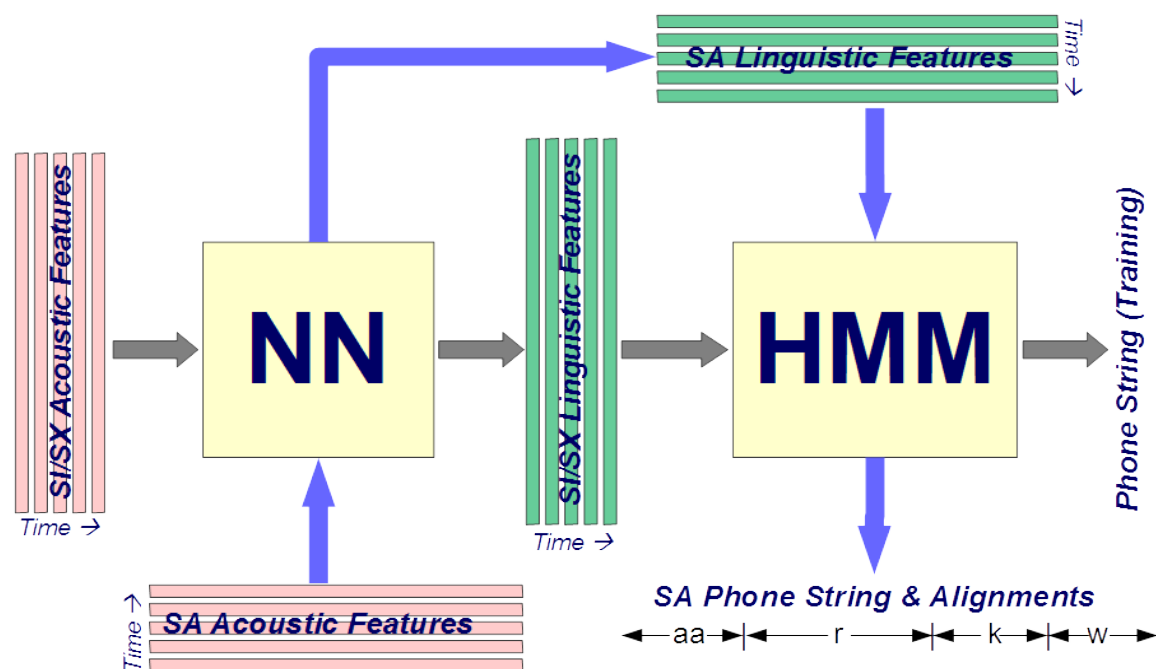
Figure 3. High-level depiction of the system's Front-End (FE) components.

---

2   The speech recognition community has a long history of using "phone" (or "phoneme") for what would be more accurately called a "phoneme-like, sub-word unit of modeling." Give the unwieldy nature of the more precise expression though, I'll just follow convention and use the "phones" misnomer in this paper.

Horizontally (and with gray arrows), Figure 3 depicts the model training processes, which use the TIMIT database's SI and SX sentences (with the standard train/test split). Input to the NN stage consists of sentential sequences of frames of MFCCs (Mel-frequency cepstral coefficients) – the predominant acoustic featurization used by the automatic speech recognition community. The output of the NN stage is comprised of sequences of frames of fairly standard acoustic/articulatory distinctive features (here labeled "linguistic features" and later referred to as "phonetic features"). The linguistic/phonetic feature sequences from the SI/SX sentences serve as training inputs to the HMM model which is tasked with learning to segment and classify the frame sequences into series of phones with associated start/stop times. Once this training is complete, we have no more use for the SI/SX recordings or their derived featural representations.

The next step in the FE is to put the trained NN and HMM to use, processing the two SA sentences as recorded by each of the TIMIT speakers. This is depicted vertically (and with blue arrows) in Figure 3. We retain the resulting sequences of frames of linguistic/phonetic features corresponding to the speakers' SA utterances to use in the post-FE stages of the method. Signal processing also extracts frame-by-frame pitch and loudness features (not illustrated in Figure 3). We utilize the trained HMM to force align each SA sentence's sequence of linguistic/phonetic frames to its canonical transcription (Figure 2). And then, the resulting phone alignments are used to segment the pitch, loudness, and phonetic feature frames (as illustrated in Figure 4) for the downstream processing.
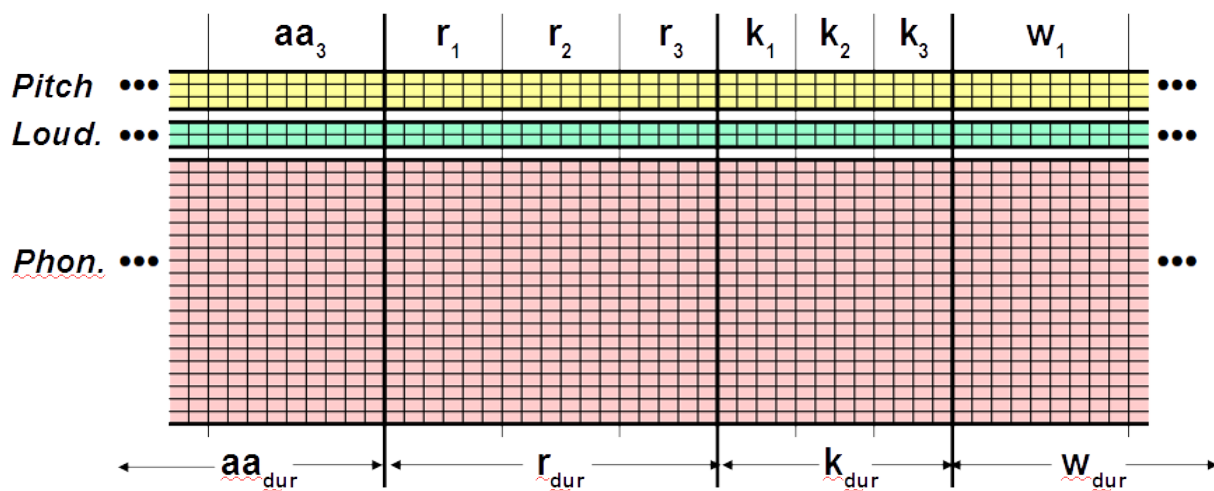
Figure 4. A partial sequence of pitch, loudness, and phonetic feature frames of one utterance, segmented into phones.

Super-Vector Selection, Projection, and Normalization – FE to BE "Glue"

Our first step in getting ready for the BE ML is to transform each SA utterance's segmented sequence of feature vectors (Figure 4) into a single super-vector (SV) for that utterance (one of the rows illustrated in Figure 5). In order to make phone internal (*e.g.*, vowel inherent spectral change (Nearey & Assmann, 1986)) and edge co-articulations available to the subsequent ML,

we temporally split each phone into thirds. Each third of each distinct phone (*e.g.*, the 6 frames corresponding to the center third of the /r/ phone, r_2, in Figure 4 above) is summarized. We, then, form an utterance SV (*Utt_SV*) by concatenating the resulting summarized representations for each phone third's pitch, loudness, and phonetic features plus duration (along with the sentence's global rate of speech, mean loudness, and mean f_0).
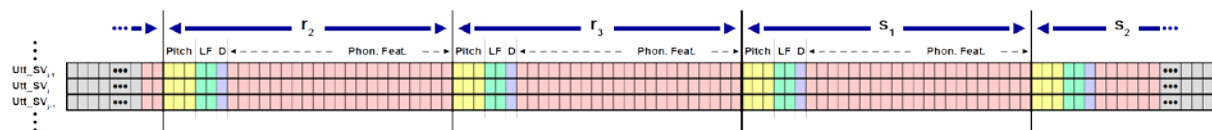

Figure 5. Partial representation of a few SA utterance super-vectors (*Utt_SV*s) derived from the segmented sequences of feature vector frames for those SA utterances.

We could, in theory, go straight to the analysis of the features of these *Utt_SV*s to determine which feature combinations are most useful in distinguishing dialects. However, there is a practical problem which must be addressed – the *Utt_SV*s are very long (on the order of 2000 elements) which makes exploration of the combinations not computationally feasible. The adopted solution to this problem was to define *partitionings* of the *Utt_SVs* into meaningful groups of features. Figure 6 illustrates such a *partitioning*, splitting the phonetic features of an *Utt_SV* into 4 groups – the thirds of the phone /r/ (*i.e.*, r_1, r_2, & r_3) plus a background group of all of the phonetic features which are not from the phone (*i.e.*, ¬r). We can then define a *partitioning series* which consists of a similar partitioning for each of the phones of a particular SA sentence. Comparisons can be made across the elements of the partitioning series since each phone-specific partitioning partitions the same global set of features from the super-vectors. Such comparisons are the basis of the graphs that we will look at in the *Results* section below.

Within a partitioning (such as that of Figure 6) we exhaustively explore each combination of feature groups, where a particular combination is conveniently represented as a group inclusion bit vector (*e.g.*, 1010 in Figure 6 represents the "not /r/" features plus the features from the middle third of /r/). For each such combination of the partitioning's groups, we select the features from each full *Utt_SV*$_j$ to create a corresponding selected utterance SV (*Utt_Sel_SV*$_j$) as illustrated in Figure 7. The resulting collection of *Utt_Sel_SV*s represents the particular subset of the features from the original super-vectors which the BE ML will have available to it for training and testing for the specified combination of feature groups (*e.g.*, 1010) within the current partitioning.

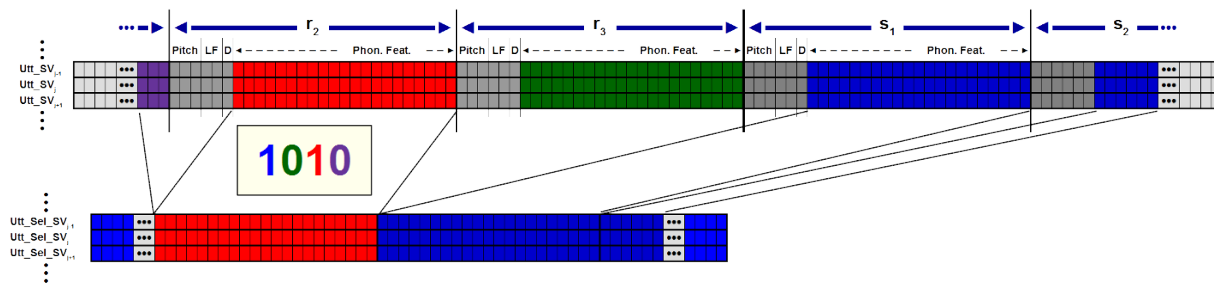| Phonetic Features				
¬r	r_3	r_2	r_1	
0	0	0	0	Baseline1
0	0	0	1	
0	0	1	0	
0	0	1	1	
0	1	0	0	
0	1	0	1	
0	1	1	0	
0	1	1	1	Topline2
1	0	0	0	Baseline2
1	0	0	1	
1	0	1	0	
1	0	1	1	
1	1	0	0	
1	1	0	1	
1	1	1	0	
1	1	1	1	Topline1

Figure 6. Partitioning.

Figure 7. Selection of feature groups for training corpus creation (resulting in the *Utt_Sel_SVs*).

The variation in the number of features selected (the vector dimension of the *Utt_Sel_SVs*) with different group selection specifications (*e.g.*, 0001 vs. 1111) is still problematic with respect to comparison of results across training conditions. For comparability, we would like the models in the BE ML to have the same number of parameters. This final "glue" issue was resolved by calculating an affine transformation matrix, via LDA (linear discriminate analysis), and applying that matrix to project each *Utt_Sel_SV*$_j$, into a very short normalized feature vector, *Utt_Norm_SV*$_j$. By producing such a collection of standard length feature vectors for each group combination within each partitioning across the partitioning series, we create conditions for fair comparisons since each NN training/evaluation in the BE ML employs the same (small) NN architecture (with equivalent modeling power). This is crucial to the feature ranking process.

Back-End Machine Learning and Evaluation

The BE ML consists of training hundreds of multi-layer perceptron (MLP) NNs – one for each feature group combination within each partitioning across the entire partitioning series – and evaluating the classification accuracy for each resulting NN for each of the DRs. For each such training, the original, full super-vectors (the *Utt_SVs*) are selected, projected, and normalized as described above to create corresponding *Utt_Norm_SVs* specific to the desired information subset.

Within each partitioning, the DR-specific accuracies are used to calculate a ranking metric score for each (non-background) feature group of the partitioning. This is done for each dialect region (DR). The process is repeated for all partitionings in the partitioning series, collecting the ranking metric scores into DR-specific tables of scores representing all of the (non-background) feature groups from across the entire partitioning series. Then, because the method has been designed to allow fair comparison across partitionings, we simply take the highest ranked feature groups within each DR table as the salient aspects of that DR's accent. We'll take a look at some examples in the results section below, but first, in order to understand those graphs, we need to briefly examine the ranking metric calculation.

Ranking Metric

The ranking metric is designed to enable fair comparison between the different (non-background) feature groups of a partitioning (*e.g.*, r$_1$, r$_2$, and r$_3$ of Figure 6) and, also, fair comparison of groups from the various partitionings of a partitioning series. It is a weighted average of five indicators of a feature group's importance with respect to identifying a DR. The

graphs of the *Results* section below show each of those five indicators (*db1%+*, *db2%+*, *dt1%-*, *dt2%-*, and *dm%+*) as bars in addition to the final ranking metric (*RMetric*) for each graphed feature group. Each indicator represents a normalized change in accuracy versus distinct references. Though not previously discussed, four of those references are marked on the right side of Figure 6 above. Those references represent feature selections where no information (0000) is used in training (*Baseline1*), only the background ("non-X") features (1000) are used (*Baseline2*), all of the partitioning's information (1111) is available for training (*Topline1*), and all of the features except the background features (0111) are used (*Topline2*).

Indicator calculations are for a given feature group of a partitioning – for example, the first third of the /r/ phone (r_1) of the partitioning of Figure 6 (*i.e.*, the 2nd bit in the feature group combination bit vector, x**1**xx). The *db1%+* and *db2%+* indicators represent deltas which we expect to be positive as we add the (r_1) group's features to a reference which does not include them. The indicator *db1%+* is the normalized increase in classification accuracy obtained when the group's features are added to the *Baseline1* features (0000→0100). The indicator *db2%+* is similar except it is the increase versus the *Baseline2* features (1000→1100).

The indicators *dt1%-* and *dt2%-* are analogous to *db1%+* and *db2%+*, but they represent deltas which we expect to be negative as we remove the (r_1) group's features from a reference which includes them. The indicator *dt1%-* is the normalized decrease in classification accuracy obtained when the group's features are removed from the *Topline1* features (1111→1011). The indicator *dt2%-* is similar except it is the decrease versus the *Topline2* features (0111→0011).

And, the final indicator *dm%+* (the mean normalized delta increase) is the average increase in classification accuracy obtained by adding the (r_1) feature group into each combination of features which do not already include it. In order not to double count the other indicators, combinations which involve *Baseline1/2* or *Topline1/2* are excluded. In our example using the first third of /r/, the delta accuracy increases from 0001→0101, 0010→0110, 1001→1101, and 1010→1110 would be included in the *dm%+* average.

RESULTS

Returning to the titular question regarding what characteristics occurring in a Bostonian's speech make it readily identifiable as being from Boston, and likewise what aspects of a Texan's speech make it identifiably Texan, we'll take a brief look at a couple of example results from applying the above procedure. It should be noted that, though these are real results, they should be regarded merely as selected illustrations of the kinds of results that one might obtain across the board as the methodology is further refined.

The first example ranking (Figure 8) is with respect to speech "from Boston," where we're generously letting TIMIT's New England DR stand in for Boston. This example is drawn from a partitioning series over the phones of the SA1 sentence using only the phonetic features. Each partitioning was into feature groups X_1, X_2, & X_3 (temporal thirds of X) plus ¬X (as discussed above), where X represents an SA1 phone. Furthermore, in this case, the components of the ranking metric for each third of X were averaged (*e.g.*, *db2%+* for /r/ is the mean of the *db2%+* values for r_1, r_2, and r_3), so that each phone is considered as a whole.
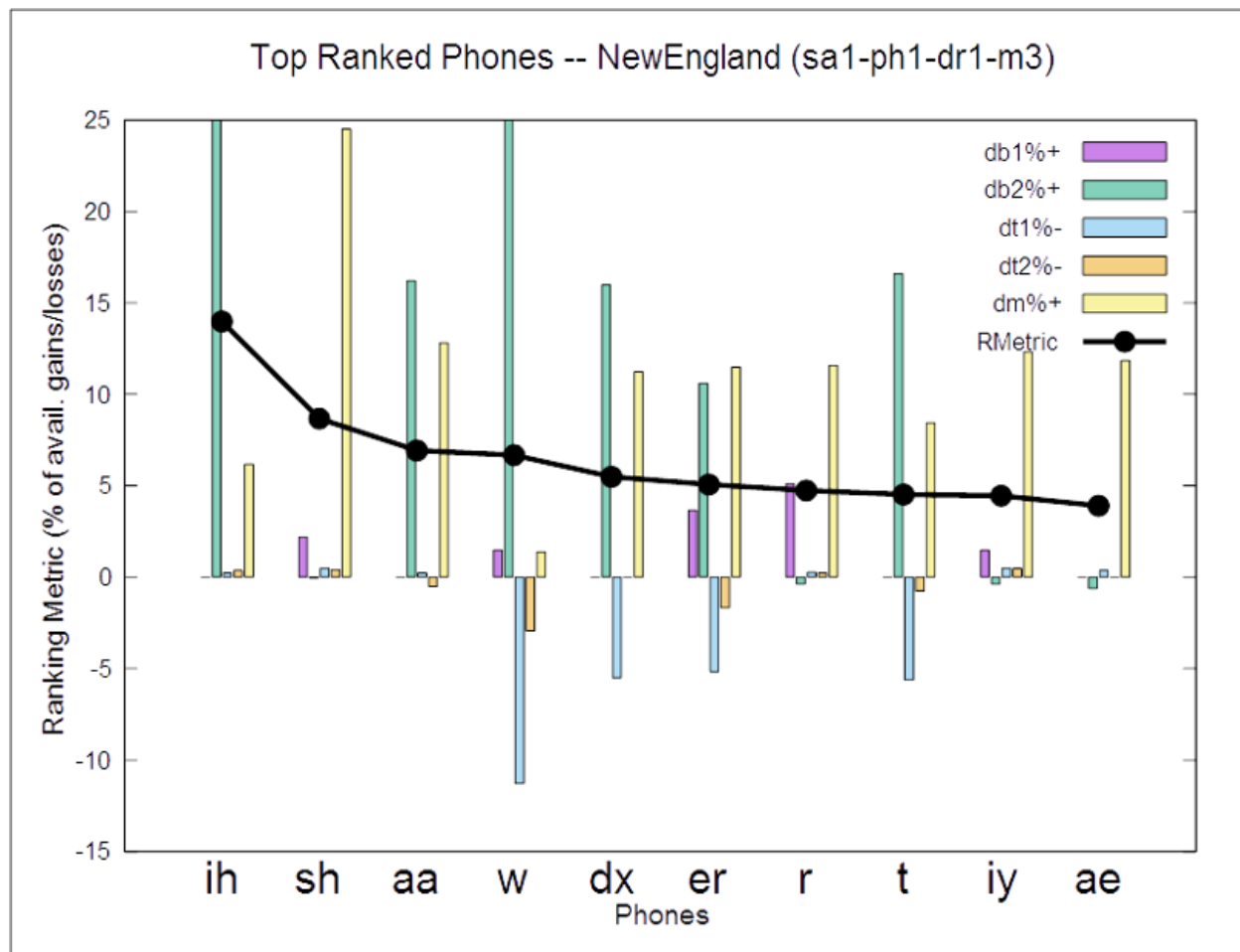
Figure 8. Example "Bostonian" results.

In a prototypical heavy Boston accent, speakers delete /r/s (*e.g.*, "Pahk duh cah in duh yahd."). So a word like "harbor" (with a canonical [ARPAbet] pronunciation of /hh aa r b er/) has a very distinctive Boston pronunciation. Looking at the Top 10 list of significant phones for identifying the New England pronunciation, we see that it includes all of those distinctive (vocalic & rhotic) phones of "harbor" (/aa/, /r/, & /er/). Also, the only occurrence of /ih/ in SA1 is juxtaposed with /er/ in the word "year" (canonically /y ih er/). We see, again, that the expected Boston (non-SAE) pronunciations of /ih/ and /er/ were automatically flagged.

Our second example ranking (Figure 9) is for speech "from Texas," where we're letting TIMIT's Southern DR represent Texas. This example is drawn from a partitioning series over the phones of the SA2 sentence using only phonetic and duration features. Each partitioning was into feature groups X_1, X_2, & X_3 (temporal thirds of X), plus X_{dur} and ¬X, where X represents an SA2 phone. Unlike in the first example, here, the phone's duration and the phone thirds were ranked independently. Figure 9 shows the Top 15 feature groups (from this larger set).
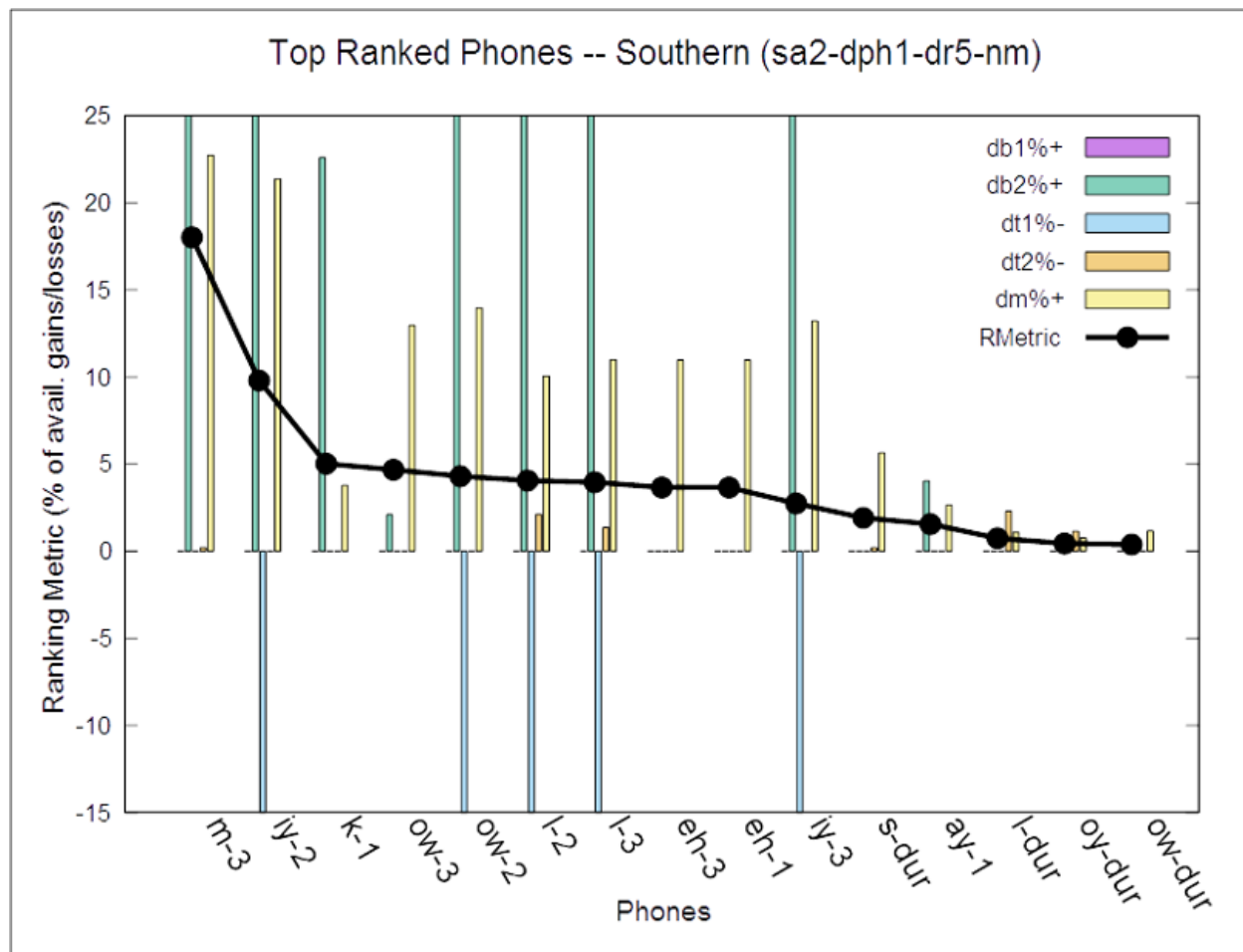
Figure 9. Example "Texan" results.

The Southern accent's vowels are known to differ from SAE in a variety of ways (Allbritten, 2011). It famously reduces the diphthong /ay/ to the monophthong /aa/, and its "drawl" embellishes other vowels with additional movement (*e.g.*, /iy/→/iy ah/ and /ow/→/ow ah/). We would, therefore, expect Southern speech to be distinguishable by its diphthongs, and indeed, all of the diphthongs which actually occur in SA2 (/ow/, /iy/, /oy/, & /ay/) are in our Top 15 list. Southern speech also shifts /eh/ into the space occupied by /ih/ in SAE (*e.g.*, "get"→"git"). We see that the Southern /eh/ is flagged. And, finally, the word "oily" (in SAE, /oy l iy/), which occurs in the SA2 sentence, has a non-standard Southern pronunciation (realized with a dark /l/ and modified diphthongs). All three phones of "oily" show up in the Top 15 distinguishing feature groups list.

DISCUSSION

The work presented in this paper represents a first cut at creating a methodology which ultimately aspires to automatically comprehensively catalog features of speech which distinguish accents of specific sub-populations. Given its provisional nature, we outlined a modest objective for the research at this stage – to provide indications that the method can be further developed into an effective technique to realize those aspirations. The example results, discussed in the

prior section, demonstrate that the method has potential. That said, those results (and others) also show evidence of spurious findings, likely due to overly powerful ML latching onto insignificant statistical regularities within the limited data. While future methodological improvements should reduce the counter-intuitive findings, there will always be some of those with a method such as this – since human and machine learners are, inherently, working on different problems with different tools at their disposal. ML merely extracts statistical regularities, exclusively based upon the limited data made available to it, while humans can't help but bring deep, interconnected knowledge to any task.

A top item on the future elaborations/improvements list is to reduce the method's modeling power. It is apparent that the combination of LDA data projection followed by multi-layer perceptron NNs in the BE allowed spurious idiosyncrasies of the data to produce better than justified accuracies. Replacing LDA (supervised) with principal components analysis (unsupervised) data projection should go a long way towards rectifying this. It would also be germane to experiment with a less powerful BE ML method such as kNN (k-Nearest Neighbor) classifiers, which would have the additional benefits of facilitating an efficient jackknifed evaluation design (also opening the door to elimination of the data projection step altogether). Fuller utilization of the available features (the current results only used the phonetic and duration features) is a priority as well.

This preliminary form of a methodology for the automated, data-driven discovery of accent discriminating acoustic features shows initial promise. Especially with the elaborations suggested above (plus numerous other improvement possibilities), there is reason to be optimistic about the prospects of evolving a viable methodology for creating useful catalogs of the characteristic acoustic aspects of sub-populations' accents. A comprehensive catalog of such (automatically-derivable) features would contribute to our explicit knowledge of the correlates of accent, but perhaps more significant would be its potential to enable the more capable and individualized CAPT tools of the future.

ACKNOWLEDGMENTS

The (open source) Kaldi speech recognition toolkit (Povey, et al., 2011) was invaluable in carrying out this work.

ABOUT THE AUTHOR

Jim Talley is the founder and CTO of Linguistic Computing Systems, an early stage start-up focusing on data-driven approaches to interesting linguistic applications. Prior to starting LingCosms, Jim worked for decades as a research scientist in industry research labs (mostly Motorola Labs [in its various incarnations] and MCC [a pre-competitive research consortium]) on speech and handwriting recognition, predictive analytics, machine learning, and human interface. Jim's graduate education was at UT Austin in Linguistics (Acoustic Phonetics), following Computer Science (and Latin American Area Studies) degrees at the University of Kansas. He was an ESL instructor in the ancient past.

REFERENCES

Derwing, T. M. & Munro, M. J. (2015). *Pronunciation Fundamentals: Evidence-based perspectives for L2 teaching and research* (Vol. 42). John Benjamins Publishing Company.

Allbritten, R. M. (2011). *Sounding Southern: Phonetic features and dialect perceptions*. Unpublished Ph.D. thesis, Georgetown University.

Garofolo, J. S., Lamel, L. F., Fisher, W. M., Fiscus, J. G., & Pallett, D. S. (1993). DARPA TIMIT acoustic-phonetic continuous speech corpus CD-ROM. NIST speech disc 1-1.1. NASA STI/Recon Technical Report N, 93, 27403.

Labov, W., Ash, S., & Boberg, C. (2005). *The atlas of North American English: Phonetics, phonology and sound change*. Walter de Gruyter.

Nearey, T. M. & Assmann, P. F. (1986). Modeling the role of inherent spectral change in vowel identification. *JASA*, 80, 1297-1308.

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.

Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., ... & Vesely, K.(2011). The Kaldi speech recognition toolkit. In *IEEE Workshop on Automatic Speech Recognition and Understanding*, 1-4.

Swan, B. & Smith, M. (Eds.) (2001). *Learner English: A teacher's guide to interference and other problems*. (2nd ed.) Cambridge Univ. Press.

Wallace, L. (2016). Using Google web speech as a springboard for identifying personal pronunciation problems. In J. Levis, H. Le, I. Lucic, E. Simpson, & S. Vo (Eds). *Proceedings of the 7th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Dallas, TX, October 2015 (pp. 180-186). Ames, IA: Iowa State University.

USING GOOGLE WEB SPEECH AS A SPRINGBOARD FOR IDENTIFYING PERSONAL PRONUNCIATION PROBLEMS

Lara Wallace, Ohio University

For students such as International Teaching Assistants who do not have heavily-accented speech but must improve their spoken English intelligibility, Google Web Speech (GWS), an ASR-based transcription tool, can be used to increase awareness of potential oral communication problems. As research has indicated, ASR (automatic speech recognition) serves as a tool to improve students' autonomy in learning, and ultimately is able to help students improve unclear pronunciation. GWS transcribes students' speech. Students then correct and mark the transcript. After analyzing the marked transcript for discourse intonation features and scrutinizing discrepancies between the transcriptions, students compare their speech to models. From this process, students may gain an understanding of where it is possible improve their delivery, including words they might not be pronouncing intelligibility. They then practice with the goal of recording an improved version. In this paper, the procedure, benefits, and limitations will be discussed.

INTRODUCTION

Students report having difficulty identifying pronunciation weaknesses; however, without knowing what they should focus on, it is challenging for them to improve their pronunciation skills effectively. One tool to help students increase awareness of potential oral communication problems and to build self-monitoring skills is to listen carefully to a recording of themselves speaking and to analyze their speech in part by marking a transcript of this audio recording (Celce-Murcia, Brinton, & Goodwin, 1996). This type of activity is encouraged in *English Communication for International Teaching Assistants* (ITAs) by Gorsuch, Meyers, Pickering, and Griffee (2013). For ITAs in particular, it is especially important that they work efficiently on improving their comprehensibility because not only do they contend with a heavy workload (teaching, studying, conducting research), they also must meet a higher standard of comprehensibility than their graduate counterparts who do not have teaching obligations. Unfortunately, ITAs who get by in their daily lives with few communication problems sometimes fail to notice when their spoken English skills result in communication breakdown, and may become discouraged when they must take a class to improve on a skill they thought they had mastered (Wallace, 2014). By listening to and analyzing an audio recording of their speech and seeing a transcript of it, ITAs can come to realize that there is room for improvement (Wallace, 2013, 2014). It is not only ITAs who can benefit from this activity, but other higher-level speakers as well.

Much of the usefulness of this activity comes from having this visual reference—a transcript of the student's speech; yet transcribing by hand is a time-consuming process and is not the

pedagogical goal. In this practitioner's experience, the quality of the transcriptions was variable and the completion rate was low when transcribing by hand was assigned as homework, and transcribing during class in order to improve quality and guarantee completion was far too time-consuming. Google Web Speech (GWS), ASR (automatic speech recognition)-based dictation software, shortens this process of transcription, taking less time to complete in class than transcribing by hand so that all students may benefit from the activity. Google Web Speech's transcription of the audio is only a beginning of students' increasing awareness. In order to gain maximum benefit, students must revise these transcriptions by correcting any word choice mistakes so that the transcription can serve as a mirror, reflecting the written form of their audio samples.

When making these corrections, it appeared that many of the discrepancies between what Google transcribed and what the students actually said illustrated some sort of pronunciation mistake. These occasional instances of low recognition could be helpful to learners. In her 2015 PSLLT presentation, Shannon McCrocklin reported on a research study that compared a fully face-to-face pronunciation workshop to one in which half of the work was completed with ASR. She found that both groups made statistically significant improvement. Although there was not a statistically significant difference between groups, the ASR group made slightly greater improvement on 5 of the 6 sounds/sound pairings investigated, indicating that ASR is a useful tool students can use to practice segmentals. Further, McCrocklin (2016) showed that introduction to ASR can significantly improve students' beliefs of their self-efficacy and autonomy in pronunciation learning. Thus, utilizing Google Web Speech as an ASR-based dictation tool not only saves time, but as McCrocklin's research indicates, it also can help students improve their pronunciation, beginning with the identification of potentially unclear pronunciation.

Using such a tool can save valuable time, thereby allowing students to focus on the analysis of the transcript so that they can discover for themselves areas they need to improve in order to speak more intelligibly. Nonetheless, teachers must realize that this tool may only be effective under two conditions: (1) that the student uses a headset microphone for maximum clarity of input, and (2) that the student's accent is not too different from the Web Speech ASR models. This practitioner has noticed that students with SPEAK Test scores of 42 and higher tend to benefit the most from these tools since the dictation tool has too much difficulty identifying what students with heavily-accented speech say. If there are too many discrepancies between the ASR-based dictation tool's transcription and what the student actually said, not only could it be an overwhelming task to identify what problems may have resulted in the miscomprehensions, but it could also be discouraging to the student, reducing the effectiveness of the learning experience. For this tool and the activity to work well, teachers and students should attempt it informally first to determine how easy each student's speech is to understand and to practice the computing skills since it requires some coordination.

PROCEDURE & PRACTICAL SUGGESTIONS

This procedure is adapted from the transcription activity in Gorsuch et al. (2013, p. 173) where students are asked to transcribe exactly what they said and how, then make corrections where needed, and practice an improved delivery. It begins by having the students simultaneously record (e.g., on Audacity) two minutes of speech while Google Web Speech transcribes it.

Once the recording and transcription are complete, students copy and paste Google Web Speech's transcription into a document twice. They keep one as it is for comparison, and correct the other so that it is an exact reflection of the audio. On the original transcript, students highlight GWS' transcription errors in another color. They should also mark in another color the discrepancy in the revised transcription. In order for the corrected transcript to more closely reflect the audio recording, students should add in any punctuation, as well as any fillers ("eh," "uh"), hesitations (…), self-correction, and false starts that they may hear.

Depending on what pronunciation target(s) the students focus on, students can also do any of the following:

- Use ALL CAPS to indicate PROMINENT words (or the stressed SYLlable of prominent words).
- Indicate thought group division by marking pauses (/).
- Insert arrows to identify change in pitch movement (↗→↘) and key choice (↑↓).
- Indicate particularly fast sections with ">>" or indicate a slowing of speech rate with "<<" (this is not covered in Gorsuch et al. 2013).

Please see Figure 1 for a sample marked transcript and the original transcription side-by-side.

Figure 1. GWS transcription and a student's revision of the transcription.

Once students have revised and marked their transcripts, they must analyze them for where they could make improvements. They can begin by looking at the highlighted discrepancies on the original transcript and listening again to the audio recording for these sections, making note of any words that seem to be mispronounced. If there is a string of words, the problem might have to do with linking, pausing, not pausing, stressing the wrong word or syllable, or there may be a grammar or word choice error. Students, perhaps with the help of their instructor or

pronunciation tutor, should try to determine what the problem might be. Apart from pronunciation, it is also valuable for students to correct any word choice or grammar errors that they notice. Again, they are encouraged to seek input from the teacher on this.

After students become aware of what they need to work on, they should begin practicing an improved delivery of the same content. Practice should be targeted and done on a small scale first. Students (perhaps with the help of the instructor or pronunciation tutor) start by isolating difficult vowels or consonants, then work on the pronunciation of these sounds in the targeted improvement of pronunciation at the word level, next the words in the context of a phrase, and so on. Practice of different word stress patterns, word choice, and grammar can be done in a similar fashion. It is important that students practice in such a way that they do not read from a script; rather, they speak extemporaneously.

One important way for students to know whether or not their speech is similar to target production is to record themselves imitating a model; they then listen again to the recording to get a sense of how similar or different their production is. To work on speech sounds, students can use University of Iowa's Sounds of Speech (http://soundsofspeech.uiowa.edu/english/english.html). To listen to the pronunciation of words in isolation, students can see a transcription and listen to an audio recording in Merriam Webster Dictionary (http://www.merriam-webster.com/), whereas for the pronunciation of words spoken in context, they can watch video clips in which the word or phrase was used (http://youglish.com/). If students have questions, they are encouraged to work with a teacher or pronunciation tutor for feedback.

Students should record their pronunciation and compare it to the models whenever possible. Even without models for comparison, students can benefit from listening to recordings of their speech for phrasal stress, prominence, thought group division, pitch movement, key choice, speech rate, and fluency. The ultimate goal with comparing their speech to a model is that students improve self-monitoring skills and practice improved pronunciation rather than repeat the same mistakes.

DISCUSSION

Benefits

Students have credited this activity with opening their eyes and their ears to what difficulties people could have understanding them (Wallace, 2013). When students see the corrected transcript of what they said, they often realize why they are asked to repeat themselves. Some comments students have made after completing this activity (when asked what they learned about their speech) include: "I pronounced the "th" sounds sometimes wrong," "I had [a] problem in pronouncing 'appearance'," "I [learned] I hesitate to speak. I think too much before speaking. And I usually repeat the words that I pronounce incorrectly."

Immediately evident to students are hesitations (…), speech rate being overly slow or fast, recasts, false starts, and fillers. Anecdotally, students have commented that when listening back to the audio to correct the GWS transcription, there were words that the students themselves had difficulty understanding. With the guidance of a teacher, students can also become more aware

of how they need to change their stress, rhythm, or intonation. As for pronunciation, Google Web Speech's transcription can often shed light on what words were unclear.

Returning to the sample transcription of Figure 1, it is possible to compare GWS' interpretation to what the student actually said. With the instructor's guidance, one can make inferences as to why the speech was not recognized, and perhaps give feedback on what can be done to improve recognition. Compare the following (Table 1):

Table 1

Comparison of GWS's Interpretation with Corrected Transcript, Student A

Examples	Google Web Speech's Incorrect Interpretation	Corrected Transcript
A	The find	Define
B	Worth	Word
C	Forest humble	For example

In examples A and B, a likely reason behind the incorrect interpretation was dentalization of the /d/. In C, it seems the speaker omitted the velar articulation of the consonant cluster, and perhaps the jaw was not open far enough. In these examples, GWS lack of recognition points to the speaker's pronunciation problems with consonants and vowels.

Looking at another student example, one can see that GWS not only points to segmental differences, but also to rhythmic differences (see Table 2).

Table 2

Comparison of GWS's Interpretation with Corrected Transcript, Student B

Examples	Google Web Speech's Incorrect Interpretation	Corrected Transcript
A	1952 2010	1950 to 2010
B	Ants y axis	And the y axis
C	The person page	The percentage
D	Lucas play score of	Look at this graph

Based on the difference in interpretation, it is plausible that in example A, the student might not have paused between "1950" and "to." In examples B and D, the student likely had difficulty

with articulating /ð/ in "the" and "this." Example C seems to show that the student did not reduce the vowel in the final syllable of "percentage." Example D seems to point to the student separating the consonant cluster /gr/. Regardless of whether or not the student or teacher can figure out reasons why GWS may not have recognized parts of a student's speech, these discrepancies between GWS' interpretation and what was actually said can shed light on what words the student pronounces differently.

Limitations

As stated earlier, when using GWS to transcribe extended speech, it is important that students' speech is not strongly accented, and that they speak into a headset microphone in order to reduce surrounding noise. A third limitation is that students should use GWS or another voice-to-speech app that does not "get to know" their voice. For this reason, it is preferable to conduct this activity in a computer lab, rather than have the students dictate something on their smart phones.

Regarding the activity itself, for maximum effectiveness, the teacher should dedicate class time to completing it. In this way, students can receive help both with the computing as well as in the analysis, correction, and practice sections (Hubbard, 2013). Of particular importance is giving students clear feedback on what they need to do to make their speech more intelligible since GWS can only indicate a different interpretation of what the students said.

CONCLUSION

Using GWS to transcribe student speech is beneficial to students in a number of ways. First, it creates a safe speaking opportunity where students can be the only ones to judge their speech. Secondly, it saves time; instead of undergoing the arduous task of transcribing audio by hand, they check the transcription while listening to their audio recordings. Also, by marking and correcting GWS' transcription, students practice their self-monitoring and analysis skills. Finally, discrepancies between GWS' interpretation of students' speech and what the students actually said can shed light on potential pronunciation problems at the word level, and sometimes with rhythm. Again, because GWS can only point to potential pronunciation problems, the instructor has the final say in what students must work on to improve their intelligibility. With these points in mind, it makes sense that an ASR dictation tool could be useful in helping students to improve their pronunciation. Future research could investigate which kinds of pronunciation errors GWS better detects and its accuracy rate in detecting those errors. Furthermore, to reduce computing and streamline the procedure, it would be helpful if an app were available that allows students to compare the original transcription to one that students can edit on screen. Taking it a step further, adding a pitch tracking option in such an app to the transcription would help students see their pitch movement and key choice since it can be difficult to hear. In this way, students would still gain the benefit of listening carefully to and analyzing and improving their speech, and the teacher's work of giving meaningful feedback would remain relevant.

ACKNOWLEDGMENTS

captures the recognition results in a text area that the students can directly edit (while preserving the original results for comparison). Thank you for comments, Jim, and thanks for the app. I would also like to thank Edna Lima and Shannon McCrocklin for their valuable input and observations on this topic.

ABOUT THE AUTHOR

Lara Wallace is a full-time lecturer in the Ohio University Linguistics Department's English Language Improvement Program where she has been teaching oral communication courses to graduate students and coordinating the Pronunciation Lab since 2005. Her research interests include CALL, pronunciation, International Teaching Assistants, and learner autonomy.

REFERENCES

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

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

Hubbard, P. (2013). Making a case for learner training in technology enhanced language learning environments. *CALICO Journal, 30*(2). http://journals.sfu.ca/CALICO/index.php/calico/article/view/945

McCrocklin, S. (2015, October). *The effectiveness of ASR-based dictation for pronunciation improvement*. Paper presented at the 7th Annual Pronunciation in Second Language Learning & Teaching Conference, Dallas, TX.

McCrocklin, S. (2016) Pronunciation learner autonomy: The potential of automatic speech recognition. *System, 57*, 25-42 doi:10.1016/j.system.2015.12.013

Wallace, L. (2013). Taking the first step: International teaching assistants' motivation to improve their spoken English intelligibility. *ITAIS Newsletter.* Retrieved from http://newsmanager.commpartners.com/tesolitais/issues/2013-08-19/1.html

Wallace, L. (2014). *International teaching assistants' (ITAs') experiences with language learning, learner autonomy, and technology as students in a requisite oral communication course*. (Electronic Dissertation). Retrieved from https://etd.ohiolink.edu/

TEACHING TIP

THE STRESS STRETCH FOR PROSODIC IMPROVEMENT IN ENGLISH WORDS AND PHRASES

Marsha J. Chan, Sunburst Media and Mission College, CA

INTRODUCTION

I've been jumping up and down in classrooms for well over 30 years to create an atmosphere in which learners of English as an additional language can find vigor, excitement, and rigor, just as my favorite high school Spanish and French teachers did in their passionate ways many decades ago. But in addition to infusing role-plays, dialogs, songs, and drama into classroom activities, I've developed several systematic techniques using movement. The Stress Stretch, which I wrote about as a recipe in New Ways in Teaching Speaking (Bailey & Savage, 1994), continues to be useful particularly for learners who have difficulty perceiving stress and intonation in spoken English. In my teaching career, most of my students come from linguistic backgrounds that are tonal and/or do not have the comparatively salient differences between stressed and unstressed syllables or long and short vowels as English. Even among relatively advanced learners, I have encountered quite a few who speak English with ease, and perhaps with general accuracy in word choice and sentence structure, but whose prosody causes confusion, delayed comprehension, misinterpretation, or misperception to varying degrees. Perhaps you, too, are familiar with learners like these.

Words are the building blocks of sentences, and words should not only be seen in written form, but heard in auditory form, and heard clearly. Heard and felt! Learning the auditory shape of a word, along with the visual shape of the word, and the meanings of the word, helps students make the word become part of their vocabulary. They can more easily recognize the words in the stream of speech and convey the proper message when they speak. In a thought group comprising multiple words, wherein the most important word receives greater vocal prominence than the others, this focus word is stressed, and its stressed syllable takes on the responsibility for conveying to the listener the most important part of the utterance. Gilbert (2008) describes this as the peak vowel in the prosody pyramid. Learners unfamiliar with the prosodic patterns of phrase level utterances in English miss-stress the utterance in various ways: they produce unnoticeable stress, or too many stressed syllables, or stresses on the wrong words and syllables. All types of improper stress can lead to miscommunication and listener discomfort.

By integrating the kinesthetic, tactile, visual, and auditory modalities of the Stress Stretch, teachers can heighten learners' perception of stressed vs. unstressed syllables and improve their production of these prosodic elements of English.

Description

The Stress Stretch is a physical activity that complements and amplifies other techniques for indicating stressed syllables (Chan, 2001). The Stress Stretch combines a physical movement with stress in words and phrases. Specifically, it requires the stretching and lowering of the body – the expansion and reduction of body height – to coincide with lexical stress or discourse prominence.[1] The Stress Stretch is useful for beginners who are acquiring the stress and intonation of English words and phrases as well as for seemingly fossilized fluent speakers of English. It can be integrated into a lesson at any level and in any language strand (pronunciation, vocabulary, grammar, listening, multi-skill).

Purposes

- To gain awareness of stressed syllables in spoken English
- To associate stress with vowel length, clarity and pitch
- To internalize these suprasegmental features into body memory
- To activate and link kinesthetic, tactile, visual and auditory learning modalities
- To pronounce polysyllabic words with proper stress and intonation

Procedure

1. For introductory lessons, prepare a list of multisyllabic words within the learners' vocabulary range. Include words of two, three, four and five syllables, as appropriate. Project the words on a screen or write them on a board. Here are some examples:
   a. Vocabulary from the New General Service List (Browne, et al, 2013)
      because develop program company another international understand responsible
   b. Academic Words (AWL) from the lecture "The Power of Music" (Chan, 2006)
      access expand classical focuses identify physical relaxation
   c. Words from the animal kingdom
      tiger elephant monkey rhinoceros penguin kangaroo eagle
   d. Words about the importance of pronunciation instruction
      pronounce specific essential communication breakdown critical indispensable implement
   e. Phrases and sentences with one prominent syllable (Chan, 2009)
      photograph album. He wants tea. What did she say? He reminded us.
2. Mark the words to indicate the stressed syllable. Here are suggested ways, with examples.
   a. Place an accent mark over the vowel in the stressed syllable.
      áccess expánd clássical cúlture fócuses idéntify phýsical relaxátion
   b. Use upper case letters for the stressed syllable.
      proNOUNCE speCIfic esSENtial communiCAtion BREAKdown CRITical indiSPENsable IMplement
   c. Underline the stressed syllable.
      tiger elephant monkey rhinoceros penguin kangaroo eagle

---

[1] In this synchronicity, the Stress Stretch is similar to the use of haptic hand gestures advocated by Acton and his colleagues (Acton, et al, 2013).

       d. Write the syllable-stress code[2]:
          because [2-2] develop [3-2] program [2-1] company [3-1] another [3-2]
          international [5-3] understand [3-3] responsible [4-2]
       e. Use boldface type on the stressed syllable.
          photograph album. He wants **tea**. What did she **say**? He re**mind**ed us.

3. Explain that a stressed syllable is generally longer, clearer, stronger, and higher in pitch than an unstressed syllable[3].
4. Pronounce each word on the list, exaggerating the qualities of the stressed syllable.
5. Sitting on a chair facing the class, model the Stress Stretch. Pronounce each word again, this time rising to a full standing position on the stressed syllable and returning to a sitting position on the unstressed syllables.
6. Have students put down their pens and books, sit up straight in their chairs, and place both feet on the ground. Have them pronounce each word, rising to a standing position on the stressed syllable and returning to a sitting position on the unstressed syllables. Ask them to feel the differences in vowel length and pitch as they repeat after you, then with you, and finally without your lead.

Advisories

1. Sit where students can see you and the words simultaneously in the same direction (e.g., in the front of the room, not in a book).
2. Encourage students to focus you (not at the text, especially in a book or on paper), observe the pronunciation (particularly stress and intonation) with their ears, eyes and tactile senses.
3. Vocalize the vowel sound from the moment you begin to rise and keep it going as you speak through the peak of the word at a higher pitch. Feel the duration of the vowel.
4. Notice the intonation (pitch) change as you move from standing to sitting and vice versa.
5. Synchronize each upward stretch only on the stressed syllable and sit on the chair for all unstressed syllables. For example, *develop:* sit-STAND-sit, or low-HIGH-low. Avoid standing on a stressed syllable and sitting on a stressed syllable.
6. Focus on primary stress; rise on only one syllable per word or phrase. Consider secondary stress as unstressed. That is, lower the body on syllables with schwa vowels, as in the second and third syllables of *focuses:* HIGH-low-low*,* as well as on syllables with clear vowels but secondary stress, as in the second syllable of *access:* HIGH-low.
7. If you teach where chairs are unavailable, crouch and stand.
8. Make accommodations for students with physical weaknesses or disabilities. Let them raise a hand, nod, or simply observe the rest of the class while pronouncing.
9. Rest between "sets" of words: show pictures, give explanations, use the words in sentences, or tell stories.

Stress Stretch Twins

When students have become comfortable doing the regular Stress Stretch, if they are mature, and if they have become comfortable interacting with each other at close range, you may have them

---

[2] The first numeral indicates the number of syllables; the second indicates the stressed syllable. (Murphy 2004)
[3] Producing a stressed syllable also requires greater pulmonary energy. (Celce-Murcia, et al, 2010)

do the Stress Stretch as twins. Find a willing student to act as your "twin" and demonstrate the technique before the whole class.

Twins: Face each other, hold hands,[4] and do the Stress Stretch together in synchrony.

Integrating the Stress Stretch into Lessons

In classes devoted to pronunciation, speaking and listening, it may be easy to justify taking 15 minutes to teach the Stress Stretch and pursue its use whenever you wish to draw attention to rhythm, stress and intonation. But what using it in other classes? Words are the building blocks of sentences, and words should not only be seen, but heard, and heard clearly. When students learn the auditory shape of a word, along with the visual shape of the word, and the meanings of the word, the word becomes part of their vocabulary.

In my experience, the Stress Stretch can be integrated into other classes as well–reading, vocabulary, grammar, writing, and multi-skill classes. After all, vocabulary is presented and used in all language classes. Here are some tips for doing so. When introducing vocabulary in a reading class, for example, draw the students' attention to one aspect at a time. Depending on the level, you may wish to spend more or less time on comprehension of the word's definition(s), its parts, spelling, and usage(s) in a sentence. Often students write notes or copy examples during this part of the lesson. Afterwards, teacher and students put down books and pens, plant feet on the floor and backsides on the edge of seats, and go through the focused listening, perception, and physical and vocal actions of the Stress Stretch. Later, when you've gone on to other parts of the lesson, and you observe a need to help students' repair improperly stressed words, start doing the Stress Stretch while you speak. With this cue, students, having had sufficient practice, will follow you, connecting the rise and fall of their bodies with the rise and fall of their voice and improving the intelligibility of their utterance. Or you may suggest, "Say that again. Do the Stress Stretch while you speak," and their body memory activates better oral production.

Viewing the Stress Stretch in Action

You can view Step 6 of the procedure described above in a classroom setting, using words ranging from one to five syllables. This ten-minute video is a recording of a live lesson from page 174 of Chan (2009) *The Stress Stretch*.

ABOUT THE AUTHOR

Marsha J. Chan is an educational consultant, 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 200 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

---

[4] Teach your students to ask permission, i.e., *May I hold hands with you for this exercise? Do you mind…? Would it be all right if…?* Allow twins to hold the opposite ends of a rolled up sheet of paper or stick as an alternative if touch is not acceptable.

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.

REFERENCES

Acton, W., Baker, A., Burri, M., & Teaman, B. (2013). Preliminaries to haptic-integrated pronunciation instruction. In. J. Levis & K. LeVelle (Eds.). *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference, Aug. 2012*. (pp. 234-244). Ames, IA: Ictibowa State University.

Browne, Charles, Brent Culligan, and Joseph Phillips (2013). A New General Service List (1.01). Retrieved from http://www.newgeneralservicelist.org/

Celce-Murcia, M., Brinton, D. M., Goodwin, J. M., & Griner, B. (2010*). Teaching pronunciation: A reference for teachers of English to speakers of other languages* (2nd ed.). (p. 131) Cambridge: Cambridge University Press

Chan, M. (1994). Stress Stretch, in *New Ways in Teaching Speaking*, Kathleen Bailey, Editor. Alexandria, VA: Teachers of English to Speakers of Other Languages.

Chan, M. (2001, March 19). *Body Movement and Drama in Language Teaching: Techniques for Teaching Stress and Intonation*. [PowerPoint slides]. Retrieved from http://www.sunburstmedia.com/present/bodymove/

Chan, M. (2006) *College Oral Communication 1*. English for Academic Success series, Byrd H., Reid, J. and Schuemann, C. Editors. National Geographic Learning (originally Boston: Houghton Mifflin Company).

Chan, M. (2009). *Phrase by Phrase Pronunciation and Listening in American English*. (p. 174). Sunnyvale: Sunburst Media.

Chan, M. [Pronunciation Doctor] (2013, October 5). *Marsha Chan's Stress Stretch*. [Video file]. Retrieved from https://www.youtube.com/watch?v=PWJv-l6OvAY

Murphy, J. M. (2004). Word-level stress in EAP oral communication. *Houghton Mifflin Academic Success Newsletter, 1(2)*, 1-5.

TEACHING TIP

FRENCH PRONUNCIATION AND VOWEL TENSION

Viviane Ruellot, Western Michigan University

INTRODUCTION

French pronunciation is characterized by considerable muscular tension (Grauberg 1997, p.148). The degree to which the muscles in articulators such as the tongue, lips, and jaws are tensed during pronunciation is particularly significant in vowel production. Indeed, unlike their English counterparts, the vowels in the French spoken in France[1] are pronounced with muscle tension that is sustained for the entire duration of the sound, resulting in significantly briefer ("me" vs. *mi*) and more stable ("bow" vs. *beau*) sounds (Tranel 1987, p. 34, Valdman 1993, p. 5).

As Valdman explains, pronunciation mistakes rarely lead to significant misunderstandings, mostly because languages are redundant enough for the other parts of the message to make up for the confusion resulting from the pronunciation mistake (1993, p. 3). However, in French, there are at least two instances in which lack of vowel tension can affect intelligibility and potentially lead to communication breakdown: in the case of the definite article of nouns that have the same feminine and masculine form (e.g., *le réceptionniste*, *la réceptionniste*), and with the third person singular direct object pronoun (e.g., *apporte-le*, *apporte-la*, bring it), especially when the context does not provide further clues. Beginner and intermediate Anglophone learners rarely pronounce the vowels in *le* and *la* as respectively a schwa and a [a] that are clearly distinct from each other. Instead, they produce a reduced version of them, which sounds like a hybrid of the two sounds. On the one hand, this pronunciation spares the speaker the obligation to take a firm position as to the gender of the noun. On the other, it can severely impact the intelligibility of the message.

Since the definite article is introduced as early as the first day of instruction, as all common French nouns are paired with an article, and since sustained vowel tension is an unfamiliar concept in English, Anglophone learners typically form the habit of pronouncing a reduced vowel for the schwa in *le* and for the [a] in *la* early on. Consequently, learners' pronunciation habits would benefit from early introduction to – and repeated review of – the concept of vowel tension.

PRESENTATION OF MUSCLE TENSION

---

[1] Relaxed vowels exist in Canadian French (Walker 1984).

One way to present muscle tension to learners is by contrasting the articulation of the French noun phrase *l'eau* (the water) with that of the English adjective "low." While *l'eau* is pronounced with sustained tension of the muscles in the mouth, cheeks, lips, and the pharynx for the entire duration of the word, muscle tension only applies to the beginning of "low," i.e., to the consonant and the first part of the vowel. The second part of the vowel is articulated with progressive relaxation of muscles, corresponding to the onset of velarization transforming the vowel into a diphthong.

As mentioned earlier, a reduced version of the vowels in *le* and *la* can significantly impair the intelligibility of a message. It is then mainly with activities focusing on the perception and pronunciation of the articles and object pronouns *le* and *la* that we can help learners better appreciate the necessity to tense their muscles when they pronounce French vowels.

PRACTICE ACTIVITIES

Perception

The first proposed activity focuses on perception practice. It can be presented to students as early as the beginner level since it features vocabulary (i.e., objects found in a classroom) introduced during the first days of instruction. Students listen to a recording in which a French native speaker mentions objects using the masculine and feminine object pronouns *le* and *la*. The students' task is to select each object referred to.

A. Une nouvelle colocataire ! (A new housemate!) You are spending a year in France on a study abroad program. You've lived with French students for a month. One of your housemates just moved out and Christine, a French student, is moving in. She asks you to help her unpack. Listen and choose between the two objects which one she asks you to bring (*apporter*) into her room, based on the word (either *le* or *la*) she used.

Christine : « S'il te plaît,
(1) apporte _____   a. le poster   b. la tablette
(2) apporte _____   a. le poster   b. la tablette
(3) apporte _____   a. le cahier   b. la règle
(4) apporte _____   a. le cahier   b. la règle
(5) apporte _____   a. le marqueur b. la trousse
(6) apporte _____   a. le marqueur b. la trousse
(7) apporte _____   a. le sac   b. la trousse
Merci ! »

193

Script: *S'il te plaît, 1. apporte-la (tablette). 2. apporte-le (poster). 3. apporte-le (cahier). 4. apporte-la (règle). 5. apporte-la (trousse). 6. apporte-le (marqueur). 7. apporte-le (sac). Merci !*

Translation: *Please, 1. bring it (tablet). 2. bring it (poster). 3. bring it (notebook). 4. bring it (ruler). 5. bring it (pencil case). 6. bring it (marker). 7. bring it (bag). Thank you!*

Perception (Student B) and production (Student A)

As mentioned earlier, it is important to regularly revisit the notion of muscle tension all along the learning process. The next activity can be introduced a few weeks after the beginning of classes, in the middle of the first college semester or the first high school year.

This activity is more challenging than the previous one, as students must pay particular attention to the article since the feminine and masculine versions of the nouns featured sound alike. After checking vocabulary comprehension, students complete the activity in pairs. It is structured in such a way that it focuses both on perception (for one student) and on production (for the other student). Each student is given a script that the other cannot see. Two scripts are provided so that the students in the pair can practice both their perception and production skills. However, students do not have to adhere to the proposed script, in the sense that they may choose the article they will use as long as they remember which one they selected, so that they can verify their partner's answers at the end of the activity.

A. Un entretien d'embauche. (A job interview.) Christine gets you an interview for a part-time job at the hotel where she works. While you wait, she discretely informs you about the persons going about the lobby. Listen to your partner read a script and decide whether Christine is referring to a man or a woman.

Regarde ! Là, c'est _____ de l'hôtel.	1. a. la propriétaire	b. le propriétaire
Derrière, c'est _____.	2. a. la réceptionniste	b. le réceptionniste
A côté, c'est _____.	3. a. la secrétaire	b. le secrétaire
Et devant, c'est _____.	4. a. la comptable	b. le comptable
La personne près de l'ascenseur, c'est _____.	5. a. la chef du personnel	b. le chef du personnel
Et là, c'est _____. Tu vas	6. a. la responsable	b. le responsable
voir : ils sont tous très sympas !	service client	service client

Script A: 1. *Regarde! Là, c'est <u>la</u> propriétaire de l'hôtel. 2. Derrière, c'est <u>le</u> réceptionniste. 3. A côté, c'est <u>le</u> secrétaire. 4. Et devant, c'est <u>la</u> comptable. 5. La personne près de l'ascenseur, c'est <u>le</u> chef du personnel. 6. Et là, c'est <u>la</u> responsable service client. Tu vas voir : ils sont tous très sympas !*

Script B: 1. *Regarde! Là, c'est le propriétaire de l'hôtel.* 2. *Derrière, c'est la réceptionniste.* 3. *A côté, c'est le secrétaire.* 4. *Et devant, c'est la comptable.* 5. *La personne près de l'ascenseur, c'est la chef du personnel.* 6. *Et là, c'est le responsable service client. Tu vas voir : ils sont tous très sympas !*

Translation: 1. *Look! Over there is the owner of the hotel.* 2. *Behind is the receptionist.* 3. *Next (to him/her) is the secretary.* 4. *And in front (of him/her) is the accountant.* 5. *The person close to the elevator is the staff manager.* 6. *And over there is the costumer services manager. You'll see: They're all very nice!*

The last activity also allows students to practice both their perception and production of the same sounds. However, it is designed for a higher proficiency level. Indeed, while the transparency of the vocabulary of the music instruments is furthered by the included visual illustrations, the remaining vocabulary and the variety of the featured structures is better suited for the end of the first college semester or high school year. This also explains the use of the target language in the context and directions.

B. Musique! Vous êtes à un concert de musique classique. La personne assise à côté de vous fait des commentaires sur les musiciens pendant qu'ils s'échauffent (*warm up*). Lisez ces commentaires pendant que votre partenaire décide si on parle d'un homme ou d'une femme.

1. Regardez comme _____ violoniste est jeune !
2. Vous voyez _____ flûtiste ? Ses doigts bougent très vite !
3. C'est pareil pour _____ pianiste. Quelle dextérité !
4. Ah, voilà _____ clarinettiste ! On dit que c'est une star de la musique classique.
5. Oh, _____ contrebassiste vient de casser une corde !
6. Hmm ! _____ trompettiste joue un peu trop fort, non ?
7. J'espère qu'on va bien entendre _____ harpiste : j'adore cet instrument !
8. Et le violoncelle aussi. J'espère que _____ violoncelliste va jouer suffisamment fort… Ah, ça va commencer…

Script : *1. Regardez comme la violoniste est jeune ! 2. Vous voyez le flûtiste ? Ses doigts bougent très vite ! 3. C'est pareil pour le pianiste. Quelle dextérité ! 4. Ah, voilà la clarinettiste ! On dit que c'est une star de la musique classique. 5. Oh, la contrebassiste vient de casser une corde ! 6. Hmm ! Le trompettiste joue un peu trop fort, non ? 7. J'espère qu'on va bien entendre le harpiste : j'adore cet instrument ! 8. Et le violoncelle aussi. J'espère que la violoncelliste va jouer suffisamment fort… Ah, ça va commencer…*

Translation*: Music! You are at a classical music concert. The person sitting next to you is making comments about the musicians while they warm up. Read these comments while your partner decides whether it is a man or a woman that is mentioned.*

Script: 1. *Look how young the violinist is! 2. Can you see the flutist? His fingers move very fast! 3. Same for the pianist. How skilled! 4. Ah, here's the clarinetist! People say she's a classical music star! 5. Oh, the stand-up base player just broke a string! 6. Hmm! The trumpet player is playing a little too loud, isn't he? 7. I hope we can hear the harp player well: I love this instrument! 8. And the cello as well. I hope the cello player plays loud enough… Ah, it's about to start…*

CONCLUSION

Vowel tension is one of the fundamental characteristics of French pronunciation, as it is part of what constitutes the articulatory basis of French (Valdman 1993, p. 5). While it rarely results in unintelligibility of the message (outside of the cases covered in this teaching tip), it can affect the comprehensibility of the message and make it difficult for the interlocutor to understand the speaker. If only for these reasons, it is worth investing learner and teacher resources in the study and practice of vowel tension in French pronunciation acquisition.

ABOUT THE AUTHOR

Viviane Ruellot is an Associate Professor of French at Western Michigan University, where she teaches French, French linguistics and applied linguistics. Her research focuses on the pedagogy and acquisition of French pronunciation by non-native speakers. She studies how feedback may help learners bridge the gap between perception and production and improve their pronunciation. She is also interested in the stages of pronunciation acquisition and the history of French pronunciation teaching.

REFERENCES

Grauberg, W. (1997). *The elements of foreign language teaching*. Clevedon, UK: Multilingual Matters

Tranel, B. (1987). *The sounds of French: An introduction*. Cambridge: Cambridge University Press.

Valdman, A. (1993). *Bien entendu! Introduction à la prononciation du Français*. Upper Saddle River, NJ: Prentice-Hall.

Walker, D. (1984). *The pronunciation of Canadian French*. Ottawa: University of Ottawa Press.

TEACHING TIP

INTELLIGIBILITY: FIVE WINNING ACTIVITIES FOR SPEAK TEST PREPARATION

Lara Wallace & Edna F. Lima, Ohio University

> Many English Language Learners (ELLs) struggle with computer-based spoken English assessments (Lowe & Yu, 2009). Apart from test anxiety and lack of confidence in their English language skills, students complain that it is difficult to speak to a computer (personal communication). In this teaching tip, we share activities that international teaching assistants (ITAs) felt helped them to speak more comprehensibly and to achieve higher scores on the SPEAK Test.

INTRODUCTION

Spoken English assessments, such as the SPEAK Test (Speaking Proficiency English Assessment Kit), are commonly used throughout universities in the United States to gauge speakers' intelligibility, of critical importance in teaching. Research on undergraduate students' perceptions of ITAs indicates that these students tend to blame their ITAs for communication breakdowns (Williams, 2011). Among factors leading to this negative perception are overall poor communicative competence (Lindemann, 2002) and poor pronunciation skills (Isaacs, 2008).

For those international teaching assistants (ITAs) whose speech is not deemed sufficiently intelligible, universities often offer services or classes designed to help them develop their speaking skills. ITAs' scores on these high-stakes tests may determine whether or not they attain or retain a teaching assistantship, often their primary source of income as graduate students. Regardless of feelings toward the SPEAK Test, many ITA educators are tasked with not only helping ITAs develop their speaking and teaching skills, but also with preparing them to score well on this exam.

Given the importance of ITA training, we have developed activities that may help our ITAs to improve their intelligibility more swiftly and to develop strategies to be more successful in the SPEAK Test. Following, we describe five activities that we have found to be helpful and would like to share with other ITA educators. These activities are *fly-swatting fillers, cell phone persuasion (*for prominence and intonation patterns), *body language for better intonation, Audacity and rhythm,* and *giving directions* (for stress and thought groups).

197

ACTIVITY 1: FLY-SWATTING FILLERS

Because the SPEAK Test is a high-stakes test, students commonly report some measure of test anxiety. Their nerves become apparent in the recordings, as for many people, their speech rate outpaces their ability to construct sentences, replaced instead with fillers, false starts, self-correction, and hesitation. In this activity, students become aware of and have the opportunity to adjust - in real time - distracting speaking habits that may interfere with their fluency or ability to communicate clearly.

Pedagogical Possibilities

The purpose of this activity is for students to build self-awareness of distracting speaking habits such as fillers, false starts, hesitation, low volume, and monotone speech. These habits do not make for clear delivery when teaching; consequently, it seems that examinees' scores are lower if their speech is rife with hesitations or fillers and false starts, or if they speak quietly with a narrow pitch range, even if they have clear pronunciation. Fillers are characterized by sounds like "um, uh, eh, nn" that often function as a way to hold the floor in conversation, letting listeners know that the speaker has more to say. False starts are instances where the speaker repeats a word or part of a word; these also seem to function as floor holders, and may be indicative of self-correction. Hesitations refer to long moments of silence, perhaps mid thought group. Too many hesitations can leave the listener with the impression that the speaker is unprepared or has trouble articulating his or her thoughts. Low volume can also be problematic, not only because it may be difficult to hear the speaker, but also because the speaker's pitch range can narrow when speaking quietly, removing valuable cues to important information, the organization of one's speech, and so on. It is good use of class time to play audio samples with these features so that students practice identifying them as a group before they need to give each other feedback.

After this warm up, students act as mirrors for each other, giving feedback in order to build their self-monitoring skills in real time. For many students, immediate feedback plus slowing their speech rate, lengthening key words, and pausing between sentences help them to reduce fillers, false starts, and long hesitations. Immediate feedback on low volume and monotone speech also helps students speak more loudly and with a broader intonation range. By the end of the semester, students should already be aware of any of these habits they may have, especially by listening back and analyzing their own audio recordings. This activity will make them aware of whether or not they still have these habits in real time as they speak.

Set Up and Directions

Students are divided into groups of 3 or 4. They take turns reading changes to a schedule (see Figure 1). The audience listens and reacts. There should be one fly swatter (or magic fairy wand) per group. Fly swatters are usually available seasonally in dollar stores and in larger grocery stores. An alternative to the fly swatter is the seemingly ubiquitous magic fairy wand—found practically all year round in dollar stores and other places that sell children's toys. In the absence of these props, students can instead tap the desk with their hands, although this probably will not leave as lasting a memory.

The speaker will go through changes to the schedule (Figure 1) in 90 seconds, and must adjust their delivery to make the audience happy. Listeners will do the following as needed:

- o Fillers, false starts, hesitation: raise the fly swatter each time to mirror this
- o Monotone: pretend to fall asleep
- o Low volume: put your hand to your ear, gesturing that you cannot hear the speaker
- o Good delivery: look content, smile

Announcement 5

Figure 1. Changes to a schedule prompt (Papajohn, 2009, p. 300).

Depending on how well students perform in the warm up, they can either keep the same roles throughout the activity, switching only when it is their turn to speak, or they can change roles with each speaker. What seems to work best is to have one speaker, one person who listens for fillers, false starts, and hesitations, and have the rest listen for monotone and low volume.

A variation of this activity is for students to show looks of confusion when the speaker does not announce the changes clearly enough (if they end on the old information and do not stress the new information, for example). In that way, the speaker can remember to

repeat and emphasize the new information, or perhaps make a comment on the change. This could be a separate role, or it could be the additional responsibility of the person who is listening for monotone speech and low volume.

Depending on the time allocated for the activity, after each speaker has a turn, or after each group member has had a turn as speaker, the group should discuss what they learned and how they can apply it to speaking more clearly in preparation for the SPEAK Test.

Table 1

Fly Swatting Fillers: The Activity in Brief

Time of Activity	15-20 minutes, including the review
Materials	- Fly swatters (or fairy wands)—one for every three or four students - Schedule prompt
Pedagogical Goals	- Build self-awareness of distracting speaking habits such as fillers, false starts, hesitation, low volume, and monotone. - In real time, work to change distracting speaking habits.
Directions	- Speaker will go through changes to the schedule, and must adjust their delivery to make the audience happy. Listeners will do the following: - Fillers, false starts, hesitation: raise your fly swatter each time to mirror this - Monotone: pretend to fall asleep - Low volume: put your hand to your ear, gesturing that you cannot hear the speaker - Good delivery: look content

ACTIVITY 2: BODY LANGUAGE FOR BETTER INTONATION

One way to do well on the SPEAK Test is for test takers to imagine that they are actually engaging in purposeful communication with another person. Yet, the test taking atmosphere, where students sit in a lab facing an unresponsive computer and speak simultaneously, makes this task challenging, in many cases resulting in a monotone audio recording that sounds less like an attempt at communication and more like the student was simply completing a boring and/or stressful task. This activity is designed to spotlight the need for communication, and stresses the importance of using body language and varying one's intonation in doing so.

Pedagogical Possibilities

In this task, the focus is on using facial expressions, hand gestures, and posture to bring out intonation variation when making recordings. With facial expressions, we review and practice smiling when happy or amused (e.g., when recommending a place to visit, talking about something we enjoy, mentioning something amusing in a story, greeting a group, etc.), and we open the eyes wider, raising the eyebrows when stressing key information (e.g., when giving directions on a map, mentioning key points in a graph, noting changes to a schedule, and so on). We review and practice hand gestures for emphasis, for comparing size, counting and more (e.g., giving directions on a map, comparing and contrasting differing ideas, explaining a graph, and so on). In terms of posture, we practice sitting up straight or leaning back instead of letting the shoulders slouch forward. We give students permission to take up space and to gesticulate. This open posture not only benefits the students by giving them better breath support, but it also gives them the appearance of confidence or of being relaxed.

In terms of intonation variation, we review three different levels: word level with prominence (stressing new information, key information, and contrasting information), "sentence" level with pitch movement (differentiating between questions and statements, indicating level of certainty, assuming listener's knowledge, and indicating whether or not the speaker is finished with the utterance- especially when listing), and "paragraph" level with key choice (for organization and emotion). By the time we review these concepts in preparation for the SPEAK Test, they have already been introduced, examined, and practiced extensively. In reviewing them, it is useful to draw analogies between what we do with pitch when speaking to what it is like on a piano keyboard (http://virtualpiano.net/) and to illustrate pitch contours with Praat so that students can more easily see and hear it in action. Incorporating body movements (such as nodding or raising the eyebrows on prominent words) may also help students anchor the speech patterns with movement. You and your students may find it interesting to compare recordings made with neutral body language before the review, then with animated body language after the activity, and view the pitch contours through Praat to note any differences in pitch range.

Set Up and Directions

Ideally, this activity is done in a computer lab so that students can connect with each other in pairs via Skype, Google Hangouts, or Facetime in order to better simulate the testing atmosphere. If this is not possible, students pair up and sit face-to-face. Either way, they need to be able to sit back far enough to see each other's hands and faces in order to mirror each other's body language.

Students take turns asking each other the questions below based on the map (Figure 2). During the 60-second response, the student who is listening mirrors the speaker's body language and pays attention to the speaker's intonation variation. If the speaker is monotone, the listener should make exaggerated gestures and facial expressions until the speaker catches on (if the speaker does not catch on, the listener can tell them what they

need to do instead). Afterwards, it is helpful to have the students share their experiences with this activity; specifically, what they noticed about body language and intonation, as well as how they can apply what they have learned when taking the SPEAK Test.

- o I will get to your town an hour before you can meet me. Could you recommend something I could do while I wait?
- o I'm an avid reader but I do not want to look at my mobile device. Could you recommend a place I could go to find something to read?
- o I'd like to see a movie. Could you recommend something for me to rent at the video store?
- o I love to spend time outside. What do you recommend I do at the Forest Preserve?

Table 2

Body Language for Better Intonation: The Activity in Brief

Time of Activity	15-20 minutes, including the review
Materials	- Speaking prompts (If possible) a computer lab with Skype, Face Time, or Google Hangouts
Pedagogical Goals	- Using and being aware of use of body language (facial expressions, hand gestures, posture) - Prominence: for new information, key information, and contrasting information - Pitch movement: for questions or statements, assuming listener's knowledge, and more - Key choice: for organization and emotion
Directions	- Student A asks a question based on the map (see speaking prompts). - Student B answers (60 seconds), while Student A mirrors B's body language (facial expressions, hand gestures, and posture). If B is monotone or low energy, A makes exaggerated gestures and facial expressions until B catches on (if B does not catch on, A can tell them what they need to do instead). - Students switch roles.

ACTIVITY 3: CELL PHONE PERSUASION

Similar to Activity 2, *body language for better intonation*, this activity also encourages students to use body language in order to broaden their pitch range. Because it is common practice for people to walk around talking on cell phones, students can do this activity just about anywhere without feeling self-conscious. This activity serves, in part, to make

the bridge between talking to someone on the phone and talking to no one, thus providing students with mindful speaking opportunities that they can utilize on their own.

Pedagogical Possibilities

In this task, we focus on prominence and intonation patterns. For that purpose, we use a "persuasion" prompt from the SPEAK Test (e.g., convince your roommate not to smoke in your apartment), which requires students to use intonation (including prominence, pitch variation, and intonation patterns) and body language to convince someone to do what they request. By this point, students have a clear and demonstrated understanding of these pronunciation features.

To review these features, students are given examples on how to use prominence to emphasize key words and on how to use intonation patterns (rising and falling) to convey their message successfully. They are encouraged to write down key words that they will use in the recording and practice those words for a few minutes. If they are uncertain about the pronunciation of a word, we encourage them to use Merriam-Webster online (http://www.merriam-webster.com/) in order to listen to a model and look at the phonetic transcription, including primary stress placement.

Setup and Directions

Before students receive the prompt, we explain the purpose of the activity and give them instructions (projected on the screen throughout the activity) for completing it. Then, students are given the following prompt: Imagine that you are talking to a friend who needs to pass the SPEAK Test, but has done little to improve his/her English. Persuade them to work on their English, and give them a few recommendations on how they can do so effectively.

Instead of recording this on a computer, students are instructed to answer the prompt in the form of an imaginary phone conversation, in this case, with their friend. While making this 45 to 60-second recording, students are told they may get up and walk around, and that they should make hand gestures and facial expressions that they normally would use when trying to persuade someone to do as they wish. Once they finish recording, they exchange phones with a partner, listening and providing feedback on the quality of their partner's overall response and on the specific targets: prominence and intonation patterns. Students may re-record an improved version, if necessary.

For students who do not have smart phones, we loan them a portable recording device or allow them to use one of the computers in the lab. However, if they are to use a computer in the lab, they must stand up and record their answers to the prompt moving their body appropriately. The key element here is to have students move their bodies while recording their response in order to elicit more natural speech (intonation and body language).

Table 3

Cell Phone Persuasion: The Activity in Brief

Time of Activity	20-30 minutes, including the review
Materials	● A prompt ● A pen or pencil and paper for notes ● A cell phone or a portable recording device
Pedagogical Goals	● Work on prominence. ● Focus on intonation patterns.
Directions	● Prompt: Imagine that you are talking to a friend who needs to pass the SPEAK Test, but has done little to improve his/her English. Persuade them to work on their English, and give them a few recommendations on how they can do so effectively. ● Students write down words that they will need to say in their recording. They practice those words. ● Students go away for 5 minutes and record their answers (45-60 seconds). ● They bring their device back and switch with a partner. ● They listen to the recording and give feedback to each other on the overall answer and on the specific targets**.**

ACTIVITY 4: AUDACITY AND RHYTHM

The purpose of this activity is to help students improve their rhythm by 'imitating' a model. The main focus here is on linking and reduction. However, this activity can be adapted to target any given pronunciation feature (e.g., segmentals, word stress, prominence, intonation patterns) for a variety of purposes. Since the purpose here is to prepare for the SPEAK Test, the model provided is a sample response to a prompt on the test. This activity is conducted in a computer lab; however, if students have personal computers with Audacity installed, it also works. In fact, doing this activity in class can motivate students to try this on their own machines as well.

Pedagogical Possibilities

In this individual task, we focus on imitation to help students become aware of and practice rhythm in English (perception and production). It is also our purpose to help students develop self-monitoring skills, a key element in pronunciation improvement. This activity is often conducted in class (or assigned for homework) after students have read and watched materials on English rhythm so that they can better understand the concepts by putting them to practice. It is noteworthy that at this point the students are

already familiar with imitation tasks as well. Students' attention should be focused on the target features rather than on the novelty of the activity.

Setup and Directions

This task is conducted individually as opposed to the group and pair tasks described above. Students listen to a sample answer to a "define and explain" question that is segmented so that they can repeat each sentence after the speaker. By "segmented," we mean that silence is inserted between sentences for easier imitation (see Lima, 2015 for instructions on how to do this).

The students open the file in Audacity and record each sentence after the speaker. Then, they listen to their recording and monitor their progress. They are prompted to analyze how well they imitated the speaker, in this case, in terms of linking and reduction. They are then instructed to re-record as many times as they wish until they are satisfied with their performance. Afterwards, they record a 60-second uninterrupted response (define a term in their own field of study) in Audacity. Once again, they listen to their recordings to analyze how they did, and record again if they believe there is room for improvement.

Table 4

Audacity and Rhythm: The Activity in Brief

Time of Activity	25-35 minutes, including the review
Materials	● Computers (a computer lab if possible) ● Audacity installed on the computers ● Audio file that has been segmented and silence inserted (see Lima, 2015 for instructions on how to do this)
Pedagogical Goals	● Linking (rhythm) ● Reduction (rhythm) Other potential goals: ● Stress (word and phrasal) ● Prominence
Directions	● Students open the file in Audacity and record each sentence after the speaker. They should feel the rhythm. ● After they finish recording, they listen and monitor their progress. How well did they imitate the speaker in terms of linking and reduction? ● Students re-record if they see room for improvement. ● Now they define a term in their own field of study. They then record their response in Audacity (60 seconds).

● They listen to their recording and analyze how they did.

ACTIVITY 5: GIVING DIRECTIONS

In this activity, students practice giving directions to each other by using a map. Because so many people rely on directions GPS gives them rather than a map to find their way, providing extra practice opportunities for map reading can boost students' confidence. There are a number of sample maps available online, and for this activity, the simpler the better. The SPEAK Test typically has a map that is laid out in a grid with clear blocks, common street names, and names of shops and other places typically found in a town. The map we provide below from Papajohn's (2009) is stylistically typical of what is on the SPEAK Test.

Pedagogical Possibilities

In addition to working with map reading vocabulary and giving commands (e.g., turn right, go through the intersection, walk three blocks), this particular activity lends itself well to thought group division, as well as phrasal stress and word stress practice, especially for compound nouns since the map has many.

To give directions in a way that someone can easily understand and remember, making shorter phrases and pausing between each step is important. To illustrate, which set of directions is easier to follow (pauses are marked with /)?

Example A: Exit onto Main St and walk three blocks and turn left when you see Park Place because your destination will be on the right.

Example B: Exit onto Main St / and walk three blocks // Turn left when you see Park Place // Your destination will be on the right.

Example B should be much easier to follow. The above examples can be used to illustrate the importance of pausing. As a warm up, students can listen to a set of directions and mark the transcript for pausing (/), then read it back.

When reviewing stress, it is helpful first to have the students identify where the stress is; listening and repeating is a good way for them to feel the stress. In these examples, the stress is indicated by capital letters:

    o   the BUS station
    o   the POST office
    o   the PET store
    o   turn LEFT

- o   go STRAIGHT
- o   EXit onto MAIN street.

This is not the time to debate syllable boundaries; rather, we should focus on the vowel that is stressed. Once the students successfully identify the stressed syllables, eliciting the qualities of a stressed syllable is next. Students hear these stressed syllables as longer in length/duration, slightly louder in volume, and usually higher in pitch. Take a few minutes to focus on and emphasize each quality by accompanying it with appropriate hand gestures or other body movements, such as Marsha Chan's Stress Stretch that she shared in PSLLT 2015 and demonstrates in this video: (https://www.youtube.com/watch?v=PWJv-l6OvAY).

This is also an opportunity to practice when to use articles. The general pattern of use is that students should use "the" when referring to a place on a map (there is usually just one place of each kind on the map), and no article in front of proper nouns, such as street names or named landmarks. See Table 5 for examples.

Table 5

Article Use with Places on a Map

Map	Article Use
Proper nouns e.g., Main St., 5th Avenue, Central Lake	(none)
Other nouns e.g., the bakery, the bus station, the library, the intersection	the

Set Up and Directions

For this activity, students pair up and sit back to back; the instructor informs them that they are not to twist or turn around when talking with each other. Instead, they will need to speak loudly enough and stress the key information (directions and place names) clearly. Each student has something with which to write, and a copy of the same map (see Figure 2 below). Members of each pair will take turns giving directions from one place to another. Student A begins by asking Student B how to get from one place to another of their choosing, for example, from the train station to the coffee shop. Student B gives directions, while Student A traces or marks the route student B gives. Student A reads back the route
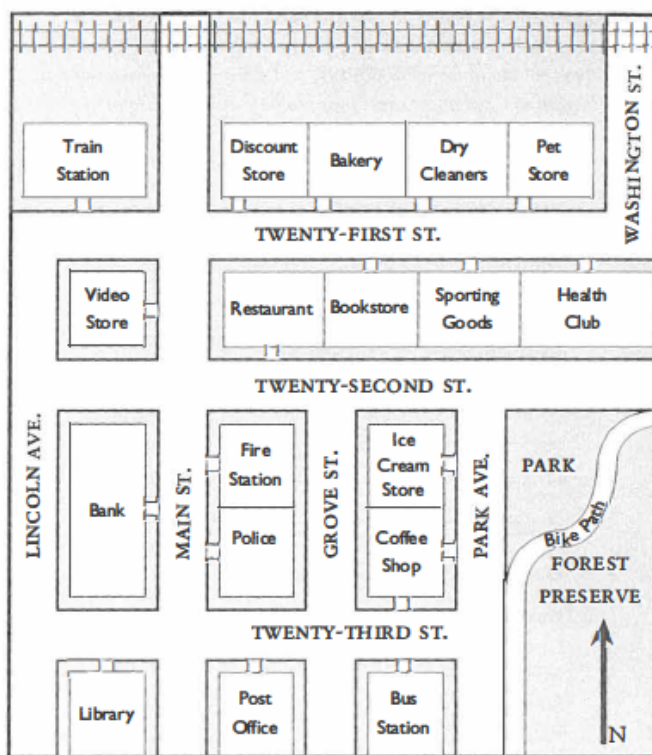
Figure 2. Sample Map (Papajohn, 2009, p. 251).

Keys to successfully completing this are to pause between each step, and stress the directions (left, right, straight, through) and the place names. Also important is the communication style. Rather than completing a task or making a recording, two activities that students might speed through, the students are practicing explaining to another person in such a way that the other person understands the directions. If their partners do not understand the directions, they will find out because either the partner will ask during the task, or when their partners read back the route, it will be incorrect.

After the activity, the teacher can ask students to share what they noticed or what they learned, particularly what was the most effective for clear communication. The class can then discuss what strategies or approaches they can take when addressing this prompt on the SPEAK Test.

Table 6

Giving Directions: The Activity in Brief

Time of Activity	20-30 minutes, including the review
Materials	- A copy of a map for each student - A pen or pencil - Chairs arranged in pairs, back to back
Pedagogical Goals	- Stress (word and phrasal) - Thought group division (pausing) - Giving directions / commands - Reading a map - Article use
Directions	- Student A asks from where to where (e.g., from the). - Student B gives directions while Student A traces the route Student B gives. - Student A reads back the route. - Student B asks from where to where (e.g., from the). - Student A gives directions while Student B traces the route Student A gives. - Student B reads back the route.

ABOUT THE AUTHORS

Dr. Lara Wallace is a lecturer and the Coordinator of ELIP's Pronunciation & Presentation Lab in the English Language Improvement Program (ELIP), a part of the Department of Linguistics at Ohio University. She has taught international teaching assistants for ten years and oral communication for 18 years. Her research interests include International Teaching Assistants, CALL, pronunciation pedagogy, student motivation for learning, and empowering education. Contact her at wallacl1@ohio.edu.

Dr. Edna Lima is a Lecturer in the English Language Improvement Program (ELIP) in the Linguistics Department at Ohio University, where she is currently teaching Oral Communication Skills for International Teaching Assistants and Writing for Research to graduate students. Edna has taught EFL/ESL for the past 20 years. She earned her Ph.D. in Applied Linguistics and Technology from Iowa State University. Her research interests include CALL, language assessment, material design, and technology applied to language learning and teaching, especially to pronunciation instruction. Contact her at lima@ohio.edu.

REFERENCES

Lima, E. F. (2015). Feel the rhythm! Fun and effective pronunciation practice using *Audacity* and sitcom scenes (Teaching Tip). In J. Levis, R. Mohamed, M. Qian & Z. Zhou (Eds). *Proceedings of the 6th Pronunciation in Second Language Learning and Teaching Conference* (ISSN 2380-9566), Santa Barbara, CA (pp. 274-281). Ames, IA: Iowa State University.

Lowe, J., & Yu, X. (2009). Computer assisted testing of spoken English: A study of the SFLEP college English oral test system in China. *Systemics, Cybernetics, and Informatics, 7*(3), 33-38.

Isaacs, T. (2008). Towards defining a valid assessment criterion of pronunciation proficiency in non-native English-speaking graduate students. *The Canadian Modern Language Review*, *64*(4), 555–580.

Lindemann, S. (2002). Listening with an attitude: A model of native-speaker comprehension of non-native speakers in the United States. *Language in Society*, *31*(3), 419–441.

Papajohn, D. (2009). *Toward speaking excellence: The Michigan guide to maximizing your performance on the TSE Test and other speaking tests* (2nd ed.). Ann Arbor, MI: The University of Michigan Press.

Williams, G. M. (2011). Examining classroom negotiation strategies of international teaching assistants. *International Journal for the Scholarship of Teaching and Learning*, *5*(1), 1–16.

TEACHING TIP

LEARNING L2 PRONUNCIATION WHILE STUDYING VOCABULARY

Nadine de Moras, Brescia University College

INTRODUCTION

Teaching practices can be more efficient if they are similar to the way the brain functions. Findings in L1 acquisition research give us important insights into successful L2 language learning, as in the normal course of events, children succeed in being proficient in their L1. According to the Usage Based Model (Ellis, 2002), frequency of items is a major contributor to the acquisition of items and structures. Once speakers have acquired a sufficient number of items with all their lexical and phonetic properties, word sequences, chunks and formulaic language in their database, they can naturally create abstract categories which enable them to apply a rule systematically, and produce novel sequences which they had never heard before.

Using this model, de Moras (2011) compared the different types of instructions (repetitions, explanations and feedback) for L2 learners; the group with repetitions had greater progress than the other groups. This study also showed that even a short period of training (30 minutes) can have a statistically significant impact on the student's learning of pronunciation.

In order to successfully register information in an internal database and be capable of using this database proficiently, learners need to have heard and practiced numerous sequences, so that the data base is complete enough for the brain to extract and extrapolate the information. Therefore, an efficient way to teach pronunciation is to concentrate on one structure, give minimal explanations with maximal practice (numerous items and sequences) until learners can use the new structure automatically.

The most frequent *obligatory liaison* is between a determiner and a noun starting with a vowel. The instructor can teach some nouns, and choose ones which are similar to the students' L1: *un_ogre, des_ogres; un_aigle, des_aigles*. This way, the student learns new words in different contexts (*l'aigle, un aigle, les aigles, des aigles*), and also intuitively extrapolates that each time *un* is followed by a vowel, an /n/ affix is attached to the initial consonant of word2, and each time *les* and *des* are followed by a vowel, a /z/ affix is attached to the initial consonant of word2. The student learns new words, affixes attached to the nouns in the singular and plural form, and the correct pronunciation, all at the same time. This is one aspect of teaching (repetition and automation) which Gatbonton & Segalowitz (2005) advocate with the ACCESS method, a Task-Based Language Teaching methodology.

For the first step of automation to be achieved, the learners need to hear word sequences (vocabulary) until they have an auditory memory of the sequences. Then, they need to practice these items and receive feedback, to counterbalance the potential influence, or interference of

their L1. Finally, after hearing and practicing a certain number of items, learners can apply a rule to novel items in communicative contexts.

If learners are asked to talk and practice a given structure without having heard it enough times, they will not have enough examples of the structure to construct an auditory memory of it, and will rely on other cues to fill in the gap, mainly their L1, which can cause transfer and eventually fossilization if the practice goes on without enough L2 input. Thus, practice without enough L2 input and repetition will not help them learn the second language accurately.

In order to learn a phonetic rule, one first needs to hear, practice and entrench in memory multiple items for each rule, along with their phonetic properties. One also needs to recognize that which constitutes a given structure, and to identify its components (Schmidt 1990, 2001); hence, there is a need to provide to the learner a brief and simple explanation on what to focus on during the lesson. It is important to limit the complexity of the sequence, and the rule to be learned, by eliminating as many other sources of difficulty as possible. I am going to illustrate these concepts by giving the example of French *liaisons*.

Background

French encourages open syllabification, and favours consonant-vowel contexts (Delattre, 1947) . Lambert-Drache (1997, pp. 12-13) states that in French there are 76 % open syllables, whereas in English there are 40 %. French also avoids hiatus (vowel-vowel contexts) which is called the "anti-hiatus constraint". This explains phenomena like elisions (*l'avion instead of le avion,*) and *liaisons (les (z)avions).*

Liaison occurs when a <u>latent</u> consonant is pronounced and attached to the following word, if it begins with a vowel or a mute *h*.

> Example: *les* (word1)_(z)*amis* (word2) [le.za.mi].

Francophones never pronounce an *obligatory liaison* consonant without attaching it to the following words. The cases of *unchained liaison* are very rare, only occur within an optional context, and are pronounced in this unusual manner mainly by people in the public sphere to produce a special effect (Encrevé, 1988). Either the liaison consonants are pronounced and linked to Word2 or they are not pronounced at all. In the case of obligatory liaisons, Majority Francophones (Majority Francophones are Francophones who live in a Francophone country or province such as Francophones living in France, Belgium, Quebec… as opposed to Minority Francophones who are Francophones who live in a non-Francophone environment such as in Ontario) typically produce over 95% of obligatory liaisons, the remaining 1-5% being caused mainly by hesitations (Ågren, 1973; Malécot, 1975).

L2 learners do not seem to acquire *liaisons* naturally. They tend to separate words and produce fewer *liaisons* than Francophones (Lauret, 2007, p. 59). Syllabic equality, syllabification, resyllabification of French, and *liaisons* are difficult to master for all L2 learners of French (Charliac & Motron, 1998, pp. 7-9), and are particularly challenging for Anglophones because of differences between the two phonetic systems.

When nonnative speakers of French pronounce differently from Francophones, they tend either not to pronounce the liaison consonant, or to pronounce it with word1, instead of pronouncing it with word2 (this is usually the most frequent error), or, to pronounce the liaison consonant improperly. There are complicated rules relative to obligatory, optional and forbidden liaisons, as well as exceptions which I will not discuss here.

Liaisons are important for several reasons. First, they often carry meaning. For example only the pronunciation of /z/ in *ils arrivent* /il.**z**a.ʁiv/ (they are arriving) distinguishes the plural from the singular /i.**la**.ʁiv/ (*he is arriving*). The pronunciation or non-pronunciation of a *liaison* can also distinguish between two homonyms: *les zéros* (*the zeros*) pronounced /le.**z**e.ʁo/ vs *les héros* (*the heroes*) pronounced /le.**é**.ʁo/.

Second, learners who do not pronounce phonemes or sequences correctly often do not recognize them when they are pronounced by native speakers (Sauders, 2007). Given that *liaisons* are present in virtually every sentence, they are essential for comprehension and comprehensibility.

Third, each mispronounced phonetic feature added to other errors may impede comprehensibility. Moreover, *liaison* errors may be caused by other initial errors or may cause errors to the neighboring phonemes or syllables. Isolated pronunciation errors often have repercussions to other items which are important for communication. For example, mispronouncing the nasal vowel of the indefinite masculine singular article *un* in *un effort* (*an effort*) does not necessarily hinder communication (this could still be problematic for comprehensibility, as the pronunciation of the /n/ makes it a feminine article instead of a masculine one). The pronunciation of the final *n*, pronounced as in English, can lead to stressing the wrong syllable in French (the first syllable of the Word2), having a pause inside a syntactic group, not producing the liaison, changing the intonation pattern and/or breaking the rhythm of the sentence. The combination of five mistakes triggered by one original mistake definitely is likely to affect comprehensibility. Now, if we add to this, the pronunciation of the mute *t* in *effort*, the five mistakes combined with this latter error almost guarantee, in my experience, that a native speaker will not understand these two words. This is why, in the end, everything becomes important, because each item influences the production of neighbouring ones, and adds to other (grammatical, lexical and phonetic) errors which hinder communication.

Teaching Tip: Teaching French *Liaisons*

Because *liaisons* partly depend on other features, they should be taught after basic intonation, stress, phonemes and link between graphemes and morphemes are introduced (ideally the first week of an introductory course). To teach *liaisons*, the instructor can begin by teaching the context consisting of the definite plural determiner + plural noun, because it is relatively easy to learn and is also very frequent.

Pronunciation can be integrated to the rest of the language lesson by teaching a feature which corresponds to the vocabulary, grammar and sentence structure. For example, the preceding *liaison* context can be taught while teaching articles. The only thing then needed in the language lesson is a sufficiently large number of examples, in order to ensure that there is enough repetition.

While starting to teach obligatory *liaisons*, the rule is simplified. The singular French definite article is *l + apostrophe* when the noun starts with a vowel (*l'avocat – the avocado*). This explanation momentarily avoids the problematic question of the gender of French inanimate nouns (*le* or *la*). The corresponding plural definite article is *les* (*les avocats – the avocadoes or the lawyers*). For example, *les avocats* is pronounced **/le. za.vo.ka/**. The *s* of the plural article is attached to word2 and pronounced /z/ at the initial of the next word.

I tried to find words which are similar to English (*l'olive - the olive, l'aigle - the eagle*), or/and words which are short (*l'auto – the car*) to facilitate recognition and retention. Also, the most unfamiliar and difficult French phonemes (nasal vowels, /y/, /ʁ/ etc.) were avoided as much as possible, even though it is nearly impossible to avoid difficult French phonemes in cognates.

The examples were chosen according to their closeness to the L1, and according to the complexity of the concept expressed by the word. It is easier to find pictures which correspond to concrete words than to more abstract ones (*abomination, accusation, accumulation*), and it also seems easier to pronounce and remember those words which are relatively simple and short.

The words, all of which are nouns, are first presented in a list in the singular and the plural forms; then a translation, followed by a phonetic transcription. The elements on which the learners concentrate are in red and in bold in the phonetic transcription, in order to draw attention to the important feature. After reading about 10 words, along with their phonetic transcriptions, the learners quickly see the rule. Another advantage of doing this exercise is that of the opportunity to show learners that they already know more words than they thought they knew.

Something else to take into consideration is the fact that learners need numerous repetitions; on the other hand it is also true that the activities should be interesting, intriguing and/or fun, without giving the impression that they are mechanical. Thus the instructor will strive to present the same information in different ways and from different points of view, so that the learners do not perceive activities as boring, or overly repetitive.

An 8-Minute Lesson

The following presents what can be done during an 8 minute lesson (and what, in fact, was done in the conference workshop). After welcoming the participants and briefly introducing myself, the workshop and the participants, I gave everyone a list of 9 words with their phonetic transcriptions and translations. The participants read the list in about 1 minute.

What follows is a list of French words with translations and phonetic transcriptions. The participants' reading of the elements enables them to familiarize themselves with the words, to see them a first time, and then to start thinking about the red letters.

avocat	l'avocat	les avocats
/a.vo.ka/	/**l**a.vo.ka/	/le.**z**a.vo.ka/
avocado	the avocado	the avocados
aigle	l'aigle	les aigles
/ɛgl/	/**l**ɛgl/	/le. **z**ɛgl/

eagle	the eagle	the eagles
auto	l'auto	les autos
/o.to/	/**l**o.to/	/le. **z**o.to/
car	the car	the cars
avion	l'avion	les avions
/a.vjɔ̃/	/**l**a.vjɔ̃/	/le. **z**a.vjɔ̃/
plane	the plane	the planes
île	l'île	les îles
/il/	/**l**il/	/le. **z**il/
island	the island	the islands
olive	l'olive	les olives
/o.liv/	/**l**o.liv/	/le. **z**o.liv/
olive	the olive	the olives
autruche	l'autruche	les autruches
/o.tʀyʃ/	/**l**o.tʀyʃ/	/le. **z**o.tʀyʃ/
ostrich	the ostrich	the ostriches
oreille	l'oreille	les oreilles
/o.ʀɛj/	/**l**o.ʀɛj/	/le. **z**o.ʀɛj/
ear	the ear	the ears
orange	l'orange	les oranges
/o.ʀɑ̃ʒ/	/**l**o.ʀɑ̃ʒ/	/le. **z**o.ʀɑ̃ʒ/
orange	the orange	the oranges

Once this introduction was finished, I then explained in one minute what we were doing: working on linking, while at the same time learning vocabulary. I explained that, as seen in the phonetic transcription, the s of *le*s, is pronounced at the beginning of the second word. I gave several examples from the sheet: /le.za.vo.ka/, /le.zo.liv/. I pointed with my finger to the bold red letters, and exaggerated the /z/: /le.**zzzzzz**o.liv/. I emphasized the fact that the syllabification does not correspond to the written words, and made a gesture showing that the *s* goes with Word2.

Each of the word groups was printed on a separate transparency. The transparency was first placed on a page with two pictures. The written words with the phonetic transcription were placed next to the picture, as shown by the next examples.

aigle   l'aigle
/ɛgl/   /**l**ɛgl/
eagle   the eagle

les aigles
/le.**z**ɛgl/
the eagles

I showed the pictures with the written words, with their phonetic transcription and their English translations. I read slowly each word, or word sequence, exaggerating the syllabification*: aigle, l'aigle, les aigles*. Next, I asked the participants to chorally repeat the same words with me (after they had already read the list of words, and had heard the three sequences once; and right after each word was pronounced, they were asked to pronounce it a second time). Pronouncing the words in unison was intended to put at ease participants who may not have remembered some words, or who were shy or hesitant.

I then proceeded to show the same pictures without the transparency. I said the words again, but without any written aid, providing only the auditory stimuli. The participants were asked to repeat the words using only the auditory stimuli. The initial aid was provided to help visual learners, as well as to make sure that they fully understood the concept. It was also reassuring for beginners to have visual support. However, learners also have to habituate themselves not to rely on visual clues, like written words, or phonetic transcriptions. They need to learn to rely also on auditory stimuli, and to be able to listen and repeat without any other help, in order to form an auditory memory. Furthermore, written words can actually be detrimental to learning pronunciation, because L2 learners tend to filter the L2 written words through the grapho-phonemic system of their L1, which often causes interference. For this reason, it is preferable to limit the use of visual cues.

Repeating what the instructor has pronounced is the step taken after one has heard the same utterances several times. If learners cannot repeat right after the instructor, they will not be able to do more complicated tasks. During the group's repetition, the instructor pays attention to all participants' productions (while easier to do in a small group, this is still possible in a larger group).

Next comes the feedback stage. When the individuals in the group repeat, the instructor can notice problems, and make general comments for everyone, or just repeat one or two problematic phonemes. Once everyone was able to repeat together correctly, I asked for volunteers to start describing the pictures (with no visual cue). Right after the volunteers' participation, the other participants were asked to describe what they saw on the pictures. By this time each person had heard the words at least 7 times, and everybody was able to say the words, especially if they were cognates (*olive*, *aigle...*). If someone was hesitating about how to pronounce a word, participants from the group or the instructor was able to help. In a classroom situation, with more time, the students would be asked to work in pairs instead.

Once students remembered an individual word with its article, and the right linking (*liaison*), they moved on to the next word. If the pictures are funny or intriguing it makes the task more interesting. Even though all of this may seem repetitive, the students do not perceive it as such, in my experience, because they need the repetition to remember the words, and there are different words, and different sequences. Also, as was the case in the workshop, the words are seen in different contexts: for example, first in the workshop, they were read silently, then they were heard and seen with a picture, accompanied by a phonetic transcription and a translation; next, they were heard again with no visual cues, following which they were repeated in groups, and finally they were said individually. After repeating the same procedure for 7-10 words (depending on conversations, remarks, feedback and time spent), the participants were now ready to apply the rule to novel sequences.

They were shown new pictures, with new words. I said the isolated singular word and the participants could see the following pictures and visual cues.

âne	**???**	**???**
/an/	**???**	**???**
donkey	**the donkey**	**the donkeys**
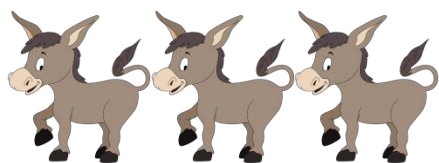
They had to come up with the right article and, most importantly, with the right pronunciation of the liaison: /le zan/; and that is exactly what the participants successfully did. They did the same with *ami* (friend) and *ogre* (ogre).They could apply the rule and pronounce the obligatory liaison with the plural determiner + plural noun context.

During this teaching tips workshop, some participants who did not know any French, and others who knew very little French, were presented with word sequences several times. After a total of 8 minutes of practice, the participants learned the singular definite article (*l'*), the plural article (*les*), 7-8 words, and were able to pronounce correctly the obligatory *liaison* after repeating and, most importantly, with novel sequences, using the correct singular and plural articles when being provided with isolated words.

CONCLUSION

At the end of the workshop, I asked the participants their thoughts on the level of difficulty of what they had practiced (from very difficult, to very easy). All participants said that what was taught was easy to learn, and one of them said that she could not believe the difference between the time when she had learned some rudimentary French in school and the learning of French in the workshop. Learners find the learning process easier if they are provided with the necessary tools: sufficient input, repetition, and practice of a limited amount of words, and one structure at a time.

The workshop participants are knowledgeable adults who have a strong phonetic background and who understand and quickly learn phonetic rules. This is not necessarily the case with average learners. Teaching less knowledgeable, dedicated students will certainly take a bit longer than it took the workshop participants. Yet, if we consider that all of the participants mastered the elements studied after 7-8 minutes, and thought that the presentation was easy to understand and remember, we can conclude that teaching pronunciation can be effective, even after a short time of instruction, as long as it is done in an efficient way.

Also, pronunciation can frequently be learned quickly by learners by virtue of the fact that it is new to them, and the novelty is attractive. Regrettably, when an incorrect pronunciation learned and practiced for over 10 years becomes fossilized, it may take much longer to unlearn and undo the wrong pronunciation, in order to learn - really learn - the right one.

ABOUT THE AUTHOR

Nadine de Moras taught ESL to Francophones in France for 10 years before teaching French to Anglophones in Canadian universities for the past 20 years. She has been teaching advanced French language and Applied Linguistics at Brescia University College (Western University affiliate) for 10 years. Her research interests include L2 pronunciation acquisition and teaching, as well as all areas of French teaching, pedagogy and material design.

REFERENCES

Ågren, J. (1973). *Étude sur quelques liaisons facultatives dans le français de conversation radiophonique: fréquences et facteurs*. Uppsala: Uppsala University Press.

Charliac, L. & Motron, A.-C (1998). *Phonétique progressive du français avec 600 exercices*. Paris : CLE International.

de Moras, N. (2011). Acquisition de la liaison et de l'enchaînement en français L2 : Le rôle de la fréquence. (Doctoral dissertation). Retrieved from Electronic Thesis and Dissertation Repository at The University of Western Ontario. Paper 176

Delattre, P. (1947). La liaison en français, tendances et classification. *The French Review*, 21 (2), 148-157.

Ellis, N.C. (2002). Frequency effects in language acquisition: A review with implication for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition 24, 143-188.*

Encrevé, P. (1988). *La liaison avec et sans enchainement. Phonologie tridimensionnelle et usages du français.* Paris: Seuil.

Gatbonton, E., & Segalowitz, N. (2005). Rethinking communicative language teaching. A focus on access on fluency. *The Canadian Modern Language Review/La Revue des langues vivantes,* 61(3), 325–353.

Lambert-Drache, M. (1997). *Sur le bout de la langue: introduction au phonétisme français.* Toronto: Canadian Scholars Press.

Lauret, B. (2007). *Enseigner la prononciation du français: questions et outils.* Paris: Hachette.

Malécot, A. (1975). French liaison as a function of grammatical, phonetic and paralinguistic variables. *Phonetica*, 32, 161-179.

Sauders, G. (2007). Vowel deletion and resyllabification in French speech: a source of perceptual difficulty for American learners of French. *Phonetics Teaching & Learning Conference.* UCL August 24-26, 2007.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics,* 11, 129-158.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge, UK: Cambridge University Press.

TEACHING TIPS

PEER-TUTORING PRONUNCIATION CONTRASTS: A FUN, EFFECTIVE CLASSROOM PROCEDURE

Lynn Henrichsen, Brigham Young University

Teachers with heterolinguistic ESL classes are sometimes reluctant to teach pronunciation because their students come from a variety of L1 backgrounds and have different pronunciation problems. Even when all students in a foreign language class speak the same L1, whole-class instruction is difficult because students are seldom equal in their ability to discriminate and/or produce contrasting English segmentals or suprasegmentals. This Teaching Tip report describes a peer-tutoring procedure that not only overcomes these challenges but actually thrives on them. In addition, it provides links to YouTube® videos that were shown in the PSLLT Teaching Tips session. Experience has shown that this procedure has many advantages: a game-like format that enlivens the classroom atmosphere, immediate feedback, individualization, automatic random sequencing, contextualization, and variation in student roles. Additionally, it provides students with persuasive evidence that correct pronunciation is important to meaningful communication.

INTRODUCTION

ESL teachers whose students speak various native languages sometimes hesitate to teach English pronunciation because these students have different pronunciation problems to overcome. For instance, an Arabic L1 speakers' usual pronunciation challenges (e.g., /p/-/b/) are quite different from those of most Japanese L1 speakers (e.g., /l/-/r/. Even when all students in a foreign language class speak the same L1, they are seldom equal in their L2 pronunciation ability. For example, some Japanese L1 speakers may have already mastered the English /l/-/r/ contrast, while other Japanese speakers still struggle with it. These differences in students' pronunciation challenges make whole-class pronunciation instruction and practice difficult because focusing on one group's pronunciation challenges leaves the other groups bored or feeling like their time is being wasted. Other problems with many traditional, whole-class pronunciation teaching and practice activities are that (1) they frequently tend to be very teacher-centered and (2) procedures that involve merely repeating after or imitating the teacher can often be boring.   As Bowen stated many years ago, "Production is an individual matter, and if meaningful guidance to students is to be offered, choral practice has serious limitations" (1972, p. 90).

This teaching tip presentation describes and demonstrates an instructional procedure that not only overcomes the challenges described above but actually thrives on them. In other words, it is especially useful when course members come from different L1 backgrounds and have different strengths and weaknesses in their pronunciations of English. Further, it involves students in

active, peer-to-peer interactions that are quite different from traditional whole-class, listen-and-repeat instruction.

Peer-tutoring activities in general have been used for decades in the development of many different skills with various types of learners (Newton, 2010). As Goodlad and Hirst (1989, p. 1) explain, "Peer tutoring is the system of instruction in which learners help each other and learn by teaching. Tutoring schemes have been used in a variety of contexts—with students teaching students, students teaching school pupils, non-professional adults teaching adults and children, and pupils teaching pupils." By their very nature, peer-tutoring activities offer an efficient and effective way of providing individualized instruction (Ehly, 1980).

Peer-interaction pronunciation-improvement activities, when used correctly, can also make the pronunciation class more student-centered, individualized, interactive, and lively. Advocated by Bowen (1975, p. 16), they were later utilized in textbooks such as Gilbert's *Clear Speech* (1993, 2001, see Figure 1) and to a lesser degree in Grant's *Well Said* (2001, see Figure 2). They are still recommended by Celce-Murcia, Brinton, and Goodwin (2010, pp. 322-323).

C

Pair practice

Student 1 says sentence (a) or (b). Student 2 says the correct answer.

1. a. He wants peas.   Not carrots?
   b. He wants peace.   Not war?

2. a. There's something in my eyes!   Call a doctor!
   b. There's something in my ice!   Call a waiter.

3. a. What does "seize" mean?   To capture.
   b. What does "cease" mean?   To stop.

4. a. Isn't this a good prize?   Yes, did you win it?
   b. Isn't this a good price?   Yes, it's really cheap.

5. a. How do you spell "trace"?   T R A C E
   b. How do you spell "trays"?   T R A Y S

6. a. Lies are terrible.   Yes, the truth is better.
   b. Lice are terrible.   Awful bugs!

7. a. How did you like the plays?   They were great!
   b. How did you like the place?   It was beautiful!

8. a. What does "Miss" mean?   An unmarried woman.
   b. What does "Ms." mean?   A woman.

Figure 1. Pair practice activity in *Clear Speech* (Gilbert, 2001, p. 46).

Figure 2. Pair practice activity in *Well Said* (Grant, 2001, p. 37).

In my instructional efforts to help English language learners improve their pronunciation, I have been using my own particular variety of peer-interaction activities for pronunciation practice (which this Teaching Tip explains) for nearly 40 years (Henrichsen, 1978; Henrichsen, 1980).

GENERAL PROCEDURE AND STUDENT/TEACHER ROLES

The particular peer-interaction pronunciation-practice procedure that I use and advocate involves putting students in groups of two (or more) based on their English pronunciation strengths and weaknesses, and then having them tutor each other on various target contrasts. A student who is

"strong" in one area (e.g., an Arabic L1 speaker who has no trouble with the /l/-/r/ contrast) tutors another student who is "weak" in that area (e.g., a Japanese L1 speaker who has trouble even hearing the difference between /l/ and /r/). Later, they can reverse roles, with the Japanese speaker tutoring the Arabic speaker on /p/-/b/, which is a difficult contrast for the Arabic speaker but not for the Japanese speaker. While students work in dyads of this sort, the teacher circulates to keep students on task, answer questions, and adjudicate when students disagree.

MATERIALS

Key to this procedure are sets of *Pronunciation Matters* peer-practice cards (Henrichsen, Green, Nishitani, & Bagley, 1999) or similar cards of your own making. The *Pronunciation Matters website* provides blackline card masters for 241 contrasting sound pairs (for vowels, consonants, consonant clusters, reduction and blending, word stress, sentence stress, intonation, and segmentation) that can be easily duplicated on a photocopy machine. Of course, independently minded teachers can make their own masters and cards.

Each set consists of twenty cards, ten that show one picture and sentence (e.g., for a word-stress contrast, a panting dog and *Look at that hot dóg*.) and ten that show a corresponding picture and sentence (e.g., a frankfurter in a bun and *Look at that hót dog*). These cards, with their pictures and minimal-pair sentences, provide structure for student tutors, as well as visual support of both the sentences' meaning and of students' progress.

LISTENING DISCRIMINATION PRACTICE

Many previous writers (Chan, 2001; Flege & Eefting, 1987; Ingram & Park, 1997; Kissling, 2014; Schneiderman, Bourdages, & Champagne, 1988) have argued that in learning new sounds in a new language, perception generally precedes production. In other words, speech perception plays an important role in the acquisition of L2 pronunciation. Therefore, in this peer-tutoring pronunciation activity, listening discrimination comes first.

In each student dyad, the student who has difficulty with the target contrast (e.g., the Arabic L1 student who cannot distinguish /p/ and /b/) works on listening discrimination first. The student for whom the target contrast is not challenging (e.g., a Japanese L1 student who has no difficulty with the /p/-/b/ contrast) serves as the tutor. The two students sit facing each other with two cards (one for each sentence in the contrasting pair) lying face-up on the desk or table between them. The tutoring student shuffles the remainder of the cards and holds them in a deck where the other student cannot see the face of the top card. The tutor then looks at that card and says (i.e., reads aloud) the sentence written on it.

The other student must then point to the card on the table that corresponds to the sentence the tutor just said. The tutor then shows the face of the card (with the picture and written sentence) to the learner. If it matches the one the learner pointed to, the tutor gives it to the learner and then goes on to say the sentence on the next card. Thus, when the learner correctly points to the card that corresponds to the sentence the tutor said, the learner is immediately rewarded by receiving the card. As the activity proceeds, the learner's increasingly large deck of cards provides continuing motivation and a feeling of growing mastery.

If the cards don't match, however, (i.e., the learner's perception was wrong), the tutor shows the learner the card and then puts it back in the deck. This "recycling" procedure results in extra practice for learners who experience difficulty perceiving the difference in the contrasting sentences. A learner may get well over 20 experiences with the 20 cards in the deck as they are recycled in this way. When the struggling learner finally receives the last card, his/her sense of accomplishment may be even greater than usual.

SPEAKING PRACTICE

After the learner receives the entire deck of practice cards via listening mode, he/she is usually ready to move into speaking mode (see Figure 3 and link to related video). In this mode, the learner follows essentially the same procedure that the tutor followed before, with the tutor pointing to the card whose sentence the learner speaks. In other words, the student who was previously the listener now becomes the speaker and tries to give the cards away. After all the cards are in the hand of the tutor, the desired level of mastery over the contrast in speaking mode is considered to have been demonstrated (although some students may want to go back and practice more). Then the two students move on to the next step.

ROLE REVERSAL

After a round of listening and then speaking practice, when the desired level of mastery has been accomplished, the two students reverse roles, and focus on a different contrast that is difficult for the student who was previously the tutor. In other words, the former learner becomes the new tutor of a different pronunciation contrast chosen because of its difficulties for the former tutor (e.g., the Arabic speaker tutors the Japanese speaker on a pronunciation challenge like word-final /r/ and /l/).

Figure 3. Students in pair practice with cards bale-veil.

COMPETITION IN TRIADS

For even more fun, students may work in triads, with one speaker and two (or more) listeners. The listener who points to the correct card first wins it (see Figure 4 as well as this video of three students competing in *bus-boss* and four students in competing *ghost-coast)*. In this case, the competition to be the first listener to point to the correct card and "win it" adds an additional, game-like element to classroom practice. In my experience, this variation can be lively and enjoyable for the participants as they compete to earn as many cards as they can.

Figure 4. Students in triads compete in listening activity with cards.

CLASSROOM ATMOSPHERE

When the whole class is practicing in pairs or triads, the classroom dynamic is radically different from that of a traditional teacher-centered, repeat-after-me pronunciation class. As Figure 5 illustrates, students are all busily working on their own, different problems time pertinent to their strengths/weaknesses in English pronunciation at the same time. The related video clip shows the desirable classroom "buzz" that results from many students busily working simultaneously on different things in the same classroom. The atmosphere is more like that in a busy workshop or laboratory, which is a welcome change from the traditional, teacher-centered, whole-class, lock-step pattern.

Figure 5. Whole class practicing different contrasts at the same time

ADVANTAGES

This procedure has many advantages, several of which have already been alluded to. First and foremost, a large class of students can work on many different pronunciation challenges that are pertinent to them individually, all at the same time. In other words, they can work on exactly the English pronunciation contrasts they find challenging and not waste time on those that are not difficult for them.

When they work as tutors, students' self esteem receives a boost, as they get to demonstrate their relative strengths in sound perception and production. This boost is comforting for students who may otherwise focus only on their pronunciation problems and begin to feel negative or discouraged.

Working in pairs, students receive immediate feedback on their pronunciation from their peer-tutor. In addition, the game-like format improves class atmosphere. Further, variation in roles as students take turns practicing listening and then speaking and alternate between being a learner and being a tutor enlivens the instructional process.

The cards—when shuffled properly—provide for automatic random sequencing in the presentation of particular practice words. Recycling the cards when learners mistakenly point to the wrong one results in a sort of automatic adjustment in the amount of practice students receive (as computer-adaptive software does) when they struggle with a difficult sound

Further, the target words are always presented in sentences on the cards. These sentences (and the accompanying complete stories, if learners are using a complete *Pronunciation Matters* unit) provide valuable contextualization that helps fill in any gaps in meaning.

Finally, the fact that the person who cannot understand a student's pronunciation of the target sound (in speaking practice mode) is a fellow student—not the "picky teacher"—provides persuasive evidence for students that (1) their pronunciation of the target sounds really is problematic and (2) correct pronunciation is important to meaningful communication.

CONCLUSION AND CAUTIONS

Although this activity involves primarily student-to-student peer-tutoring, a teacher's involvement is still very important. Sometimes student pairs come up with spurious, incorrect ways of distinguishing the two members of a target contrast (e.g., using sentence stress or intonation to distinguish a pair of utterances that actually differ only in vowel quality). Also, sometimes students do not trust the pronunciation of their student partner who is tutoring them. In such cases, the teacher must be available to intervene, check, and if necessary, set things right. In other words, while students are practicing in pairs, the teacher (and/or additional teacher aides if they are available) must constantly circulate and listen to the students as they practice with each other.

Finally, despite its many advantages, the peer-tutoring activity described in this Teaching Tip should not be used as a steady diet. As is the case with any classroom activity—even a fun one— over-use can lead to boredom. In my pronunciation classes, I use this pair-practice activity in conjunction with a variety of other whole-class pronunciation improvement procedures that focus on challenges that are more general and apply to the entire class. We usually break into pair-practice mode only once or twice a week, and usually in the latter part of a class session that has involved other instructional and practice activities. Then, students welcome the chance to work on their individual pronunciation challenges in pair-practice mode with a fellow student.

ABOUT THE AUTHOR

Lynn Henrichsen (Ed.D., University of Hawaii) teaches TESOL methods, materials development, and research methods, as well as ESL pronunciation and phonetics/phonology courses, in the Linguistics and English Language Department at Brigham Young University (Provo, Utah). During his TESOL career, he has taught, trained teachers, and/or conducted research in Mexico, Spain, Samoa, Brazil, Hawaii, Japan, Taiwan, China, Korea, Thailand, Cambodia, and Mozambique. He is the author of a dozen books and over 70 journal articles.

REFERENCES

Bowen, J. D. (1972). Contextualizing pronunciation practice in the ESOL classroom. *TESOL Quarterly, 6,* 83-94

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

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

Chan, C. (2001). The perception (and production) of English word-initial consonants by native speakers of Cantonese. *Hong Kong Journal of Applied Linguistics, 6*, 26-44.

Ehly, S. W. (1980). *Peer tutoring for individualized instruction.* Boston: Allyn and Bacon

Flege, J. E., & Eefting, W. (1987). Production and perception of English stops by native Spanish speakers. *Journal of Phonetics, 15*, 67-83.

Gilbert, J. B. (1993). *Clear speech: Pronunciation and listening comprehension in North American English* (2nd ed.). Cambridge: Cambridge University Press.

Gilbert, J. B. (2001). *Clear speech from the start: Basic pronunciation and listening comprehension in North American English.* Cambridge: Cambridge University Press.

Goodlad, S., & Hirst, B. (1989). *Peer Tutoring: A guide to learning by teaching.* London: Kogan Page.

Grant, L. (2001). *Well said: Pronunciation for clear communication.* Boston: Heinle & Heinle.

Henrichsen, L. E. (1978). Peer-tutoring pronunciation contrasts. *English Teaching Forum, 16*(2), 18-22.

Henrichsen, L. E. (1980). Peer-tutoring pronunciation contrasts. In D. L. Bouchard & L. J. Spaventa (Eds.), *A TEFL Anthology* (pp. 95-99). Washington, D. C.: International Communication Agency.

Henrichsen, L. E., Green. B. A., Nishitani, A., Bagley, C. L. (1999). *Pronunciation matters: Communicative, story-based activities for mastering the sounds of North American English.* Ann Arbor: University of Michigan Press. (Available online at www.pronunciationmatters.com)

Ingram, J. C. L., & Park, S. G. (1997). Cross-language vowel perception and production by Japanese and Korean learners of English. *Journal of Phonetics, 25,* 343-370.

Kissling, E. M. (2014). What predicts the effectiveness of foreign-language pronunciation instruction?: Investigating the role of perception and other individual differences. *The Canadian Modern Language Review/La revue canadienne des langues vivantes, 70,* 532-558.

Newton, F. B. (2010). *Students helping students: A guide for peer educators on college campuses.* San Francisco: Jossey-Bass.

Schneiderman, E., Bourdages, J., & Champagne, C. (1988). Second-language accent: The relationship between discrimination and perception in acquisition. *Language Learning, 38,* 1-19.

TEACHING TIP

INTONATION BRIDGING ACTIVITIES: MEANINGFUL PRACTICE FOR FINAL INTONATION

Greta Muller Levis, Iowa State University
John Levis, Iowa State University

INTRODUCTION

This Teaching Tip combines three variables in creating activities that both highlight this pronunciation feature (final intonation) and provide tasks demanding that language learners pay attention to something more than just the pronunciation form. The first variable is manipulation of an original task, in this case, written dialogues. The second variable involves the use of spoken language structures, especially incomplete syntactic units (e.g., Ready? Not yet.) that are commonly used in spoken conversations. The third variable includes learners' need to simultaneously decide on meaning carried by rising or falling intonation and respond to their interlocutors while both producing the appropriate language form and responding meaningfully to the other people in the dialogue.

Dialogue activities are traditionally highly controlled because they involve little more than reading aloud. As controlled activities, they also are a good first step for practicing a pronunciation feature whose meaning is most noticeable in discourse, in this case, final intonation. Dialogues by themselves can be, however, problematic for practicing intonation because of the common use of full syntactic units (e.g., yes/no questions, WH questions). These full units, especially yes/no questions with inversion, make intonational differences less important to listeners (Levis, 1999; Thompson, 1995) since questions remain questions regardless of intonation and meaning differences between rising and falling intonation are subtle rather than salient.

With short or elided sentences, meaning differences in the use of rising and falling pitch are magnified. Learners who are used to meaning being carried in the lexico-grammar of fully formed sentences are challenged to use and understand intonation which now carries grammatical meaning in the elided sentences. This demands greater creativity in production and interpretation while also focusing on intonation as a significant carrier of communicative intent.

What are bridging activities?

Bridging activities combine control and flexibility in the same activity. This is important for all language learning, but it is especially important for pronunciation because it requires both cognitive understanding and automatic motor control. Bridging activities are called "Guided Practice" by Celce-Murcia, Brinton, Goodwin and Griner (2010) and are

the middle step of their three stages of pronunciation production activities (Controlled, Guided, Communicative). Thus, bridging activities are in-between controlled and communicative practice, having some elements of both. They have enough control to allow learners to pay attention to the pronunciation topic, but distracted control because learners must also pay attention to something else such as meaning, task demands, pragmatics or other linguistic features (Figure 1).
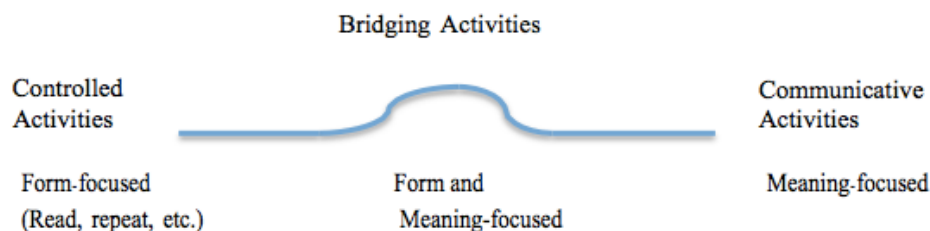
Bridging Activities

Controlled Activities		Communicative Activities
Form-focused (Read, repeat, etc.)	Form and Meaning-focused	Meaning-focused

Figure 1. Bridging activities.

Because bridging activities allow some focus on form but at the same time raise the cognitive load of the activity by forcing learners to pay attention to other issues, they are harder than controlled activities which allow full attention on pronunciation form but little on meaning. In contrast, communicative activities focus fully on the creation of meaning, but do not attend to form in general. Bridging activities may look communicative or they may look controlled, but they always allow for some, but not full, attention to form.

Final Intonation in full and short sentences

The direction of voice pitch at the ends of sentences (final intonation) communicates meaning in English, especially in interactive contexts. However, intonation is often hard to hear and its exact contribution to meaning may be hard to tease out, especially with fully formed sentences (as in the traditional dialogue below). Changing the normal dialogue exercises to elided sentences (as in the short sentence dialogue) can help learners focus on intonation and its meaning. As in any dialogue that is not created by the learners themselves, the short sentence conversational context should be discussed. For example, it is helpful to ask questions such as Who are the speakers, where is the conversation taking place, what intonation should they use, and what information do they already share. These activities are based on ideas from Allen (1971), one of the best early pedagogical treatments of intonation in TESOL, and Levis (1999).

We use full sentences and questions in normal speech, but we also use short sentences. These elided units are common in spoken language because we often do not speak in syntactically complete units. It's normal and natural in speech to do so, especially in conversations. The advantage of short sentences in teaching is both that they are a common, natural part of speech and that they highlight the contribution of intonation to the interpretation of the utterance. See the examples below.

Traditional Dialogue		Short Sentence Dialogue	
A:	Are you going?	A:	You going?
B:	Yes, but I'll go later.	B:	Later.
A:	When do you think you'll go?	A:	When?
B:	I guess I'll go around 8.	B:	Around 8.
A:	Are you driving?	A:	You driving?
B:	Yes. Do you need a ride?	B:	Need a ride?
A:	Yes, I'd like a ride, thanks.	A:	Yeah. Thanks.
B:	OK. Should I meet you here?	B:	Ok. Meet you here?
A:	No. I'll be at the library.	A:	No. At the library.

Creating short sentence bridging activities

As teachers, we are used to using and creating dialogues to teach pronunciation and a variety of other language features. For intonation, it is useful to change full-sentence dialogues into short sentence dialogues to emphasize the contribution of the intonation to the meaning. We will show four different ways to do this, but there are likely to be many others. In each case, we suggest changing just one thing: focus on memory, changing language demands, changing task demands, or increasing the attention to meaning.

Bridging Activity 1

Change your focus on memory (Short-sentence conversation)

In the first variation, we suggest changing the locus of attention from the written word to the spoken by removing the written word as a stimulus after practice. Two learners practice the dialogue several times, paying attention both to the content and to the intonation used for each utterance. After they have practiced, they should turn their papers over or the teacher's projected version of the dialogue should be turned off. Then the two learners should do the dialogue again from memory. Doing the dialogue from memory, especially if they did not know in advance to fully memorize it, creates a bridging activity in which they remember the intent of the dialogue and some or most of the lines, but may have to improvise the rest. They also have to use appropriate intonation to communicate what they want to say.

Short Sentence Conversation

Pairs. Speaker A and Speaker B read the dialogue several times using the intonations listed. Then Speakers A and B practice the dialogue again from memory, paying attention to intonation.

A: You going?   ↗

                                    B: Later.   ↘

A: When?   ↘

                                    B: Around 8.   ↘

A: You driving?   ↗

                                    B: Need a ride?   ↗

A: Yeah, thanks.   ↘

                                    B: OK, meet you here? ↗

A: OK.   ↘

Bridging Activity 2

Change the language demands (Mixed-up conversation)

In the second variation, learners can challenge their locus of attention by having to create a reasonable conversation from mixed-up lines. Because the turns involve elided sentences, the task involves more decisions and more complex language demands. First, the learners have to decide what the meaning of each utterance is likely to be and therefore what their order should be; then they have to negotiate the likely intonations that each line will use. Depending on the group of students, teachers can encourage learners <u>not</u> to write their order down, thus increasing memory demands as well. Finally, after trying out sections of the conversation, they have to try out the full conversation, using their chosen order, trying to remember how the conversation is constructed, with the intonation that is most appropriate. This makes the task far more complex than a controlled reading task.

Mixed-Up Conversation

Pairs. Work together to form a dialogue between two people using these utterances.
- The first exclamation starts the conversation.
- Decide on a sensible order of turn-taking for the rest of these utterances:
    Probably not.
    Bad?
    Cut myself.
    How?
    Bandage?
    Stupid knife!
    Not too bad.
    What?
- Practice the conversation.
- Decide what intonation to use for each short sentence.
- Perform the conversation first reading it, then perform it again from memory.

A: *Darn!*
B:
A:
B:
A:
B:
A:
B:
A:

Bridging Activity 3

Change the task demands (Build a conversation)

The next task uses a version of sentence construction activities where learners must add their own information. The first speaker uses a line from the "Openers" column, and the other speaker responds with something from the "Initial Response Words" column. The first speaker must then improvise a response (as in a normal conversation, which is co-constructed by speakers). The speakers can then start a new conversation, reversing roles. Each speaker has an opportunity to shift each conversation in one way or another. This kind of mental challenge changes the task demands. Speakers may have an idea how to start and respond with the given short sentences, but the third turn asks them to add their own response. This means that intonation choices are increasingly difficult given that speakers are focusing more on meaning. Teachers can encourage learners to use short sentences.

Build a Conversation

Pairs.
- Speaker A starts a conversation with one of the words under "Openers".
- Speaker B responds with an appropriate "Initial Response Word".
- Speaker A should then respond with their own words,
- Speakers should use rising (⤴) or falling (⤵) intonation as appropriate.
- When you are finished, start a new conversation with a different Speaker A.

Openers		Initial Response Words		Response
Coming	Going	No	Very	
Ready	Wow	Yeah	Really	(Create your own response.)
Trouble	Difficult	What	Almost	
Fun	Happy	Kind of	Later	

Example 1
Speaker A: Going? ⤴
Speaker B: Yeah. ⤵
Speaker B: Soon? ⤴

Example 2
Speaker A: Going? ⤴
Speaker B: Later. ⤵
Speaker A: OK. ⤵ See you there. ⤵

Bridging Activity 4

Change the attention to meaning (Responses)

The final bridging activity suggestion allows less planning for both people. Speaker A chooses any short sentence from the list and an intonation, using it as expressively as desired. Speaker B has to come back with a response, Speaker A has to then reply to B, and so on. The conversation should continue at least 6 turns. This kind of activity uses some control to start each conversation, but the second speaker doesn't know what's coming. So speaker B has to create a response without preparation, as does Speaker A in the next response. The conversation ends after the speakers reach at least 6 turns, and the pairs repeat the task with a new beginning. This task's attention to meaning mirrors that of free speech without long utterances. Short sentences also have to be spoken with some kind of intonation, and the brevity allows listeners to pay attention to the pitch movement as well.

Responses

Pairs.
- Pick a short sentence to start a conversation and say it with your choice of intonation (and any other expressiveness you wish).
- Your partner has to come back with any response
- Reply to their response in any appropriate way if you can.
- Continue at least 6 turns. Then start again with a new short sentence.

Example	Speaker 1:	Darn!	(short sentence)
	Speaker 2:	Again?	(response)
	Speaker 1:	I can't do this.	(response)
	Speaker 2:	What?	(response)
	Speaker 1:	The sink.	(response)
	Speaker 2:	Call a plumber?	(response)
	Speaker 2:	Not yet.	(response)

Short sentences to start a conversation

Funny?	How much?	Darn!	Ready?
Cool!	Dessert?	What for?	Shoot!

CONCLUSION

One of the most important, yet infrequent types of activities to practice pronunciation are bridging activities. Controlled activities allow complete attention to the pronunciation feature being practiced, and communicative practice allow full attention to meaning but rarely lead to correct production of challenging pronunciation forms. Bridging activities allow learners to pay attention to both form and meaning, providing challenges in producing form while attending to other elements of language at the same time. Controlled activities can usually be changed to make them less controlled and more meaningful, leading to a wide variety of bridging possibilities. The activity ideas provided here are for one of the most challenging pronunciation features, final intonation. The tasks are constrained rather than controlled; they are also meaning-oriented, fun and challenging.

ABOUT THE AUTHORS

Greta Muller Levis is a Senior Lecturer in TESL and Applied Linguistics at Iowa State University, where she also is the adviser for undergraduate students in linguistics. She has taught international teaching assistants for over 20 years. She also teaches introductory linguistics classes and second language acquisition. Her interests include the teaching of pronunciation, materials development, and effective teaching practices. She has presented papers at conferences in the United States, Czech Republic, Denmark, Canada and Taiwan.

John M. Levis is Professor of TESL and Applied Linguistics at Iowa State University. He teaches a variety of courses including Methods of Teaching Pronunciation. He is the founder of the Pronunciation in Second Language Learning and Teaching conference and has published in a variety of journals. He is a co-editor of *Social Dynamics in Second Language Accent* (DeGruyter Mouton, 2014) and of the *Handbook of English Pronunciation* (Wiley Blackwell, 2015). He is also the founding editor of the *Journal of Second Language Pronunciation* (John Benjamins).

REFERENCES

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

Celce-Murcia, M., Brinton, D., Goodwin, J., & Griner, B. (2010). T*eaching Pronunciation Hardback with Audio CDs (2): A Course Book and Reference Guide*. New York: Cambridge University Press.

Levis, J. M. (1999). Intonation in theory and practice, revisited. *TESOL Quarterly*, *33*(1), 37-63.

TEACHING TIPS

**TEACHING TALK, TELL-BACKS,
AND A DECLARATIVE TO PROCEDURAL KNOWLEDGE INTERFACE**

Marnie Reed, Boston University

Can explicit, declarative knowledge be converted to implicit, procedural knowledge? This Teaching Tip advocates the use of Teaching Talk, defined as succinct and therefore retrievable language of instruction, restated by learners as Tell-Backs for the purpose of internalizing pronunciation concepts to establish declarative knowledge. The intervening mechanism is prompted production, a form of error feedback that is achieved when the language of intervention matches the language of instruction that was used to introduce a target feature, and is in turn consistent with the tell-backs that learners use to internalize those features. Prompted production promotes self-monitoring and self-correction, and serves as an interface to bridge the declarative to procedural knowledge gap.

INTRODUCTION

Moyer (2014) identified cognitive, experiential, and psychological factors common to learners who were exceptional with respect to second language phonology. Levis (2015) investigated learners who were sufficiently proficient to engage in graduate study in English, but whose "beliefs made improvement in pronunciation difficult" (p. A42). According to Levis, the "largely fossilized" learners in his study had difficulties they did not know how to fix; that is, they "often did not have a clue how to improve" (p. A52).

This Teaching Tip takes one of the factors noted by Moyer to be common to nearly all the exceptional learners studied, a metacognitive approach to language learning (Moyer 2014, p. 7), to inform an instructional approach to improve the pronunciation of learners who do not know how to address their pronunciation deficiencies. By proposing an interface to bridge the explicit to implicit knowledge gap, this Teaching Tip promotes a metacognitive coaching approach to providing pronunciation feedback, and offers strategies for learner self-monitoring and conversion of conscious declarative knowledge to unconscious procedural knowledge.

As discovered by Derwing & Rossiter (2002) and confirmed by Foote, Holtby, & Derwing (2011), at early stages of acquisition, learners often don't know what they don't know. They are, in short, at the unconscious incompetence stage of development. They may be inadvertently mispronouncing individual segments, adding or deleting sounds in syllable onsets or codas, stressing incorrect syllables in multisyllabic words, phrases, or sentences, or misusing or entirely missing out on the pragmatic functions of intonation. As a result, even fairly fluent leaners may be unconsciously incompetent at various aspects of segmental and/or suprasegmental phonology. Instruction must advance learners beyond awareness of their erroneous productions. The stages can be visualized using a model of learner progress proposed by Reed & Michaud (2005, 2010).

Table 1

Four Levels of Competence: Achieving Unconscious Competence

		The Four Levels of Competence	
		Consciousness	*Competence*
Level 4	Unconscious Competence	−	+
Level 3	Conscious Competence	+	+
Level 2	Conscious Incompetence	+	−
Level 1	Unconscious Incompetence	−	−

Since exposure to target language input is acknowledged to be insufficient to create changes in learner output (Flege & Hillenbrand, 1984; Flege, 1993; Strange, 1995; DeKeyser, 2005), alternative candidates for achieving target-like spontaneous production are needed. Can explicit knowledge - described variably as declarative, accessible, controlled, and conscious (Bialystok, 1982; DeKeyser, 2003; N. Ellis, 2005) be converted to implicit knowledge – described as procedural, inaccessible, automatic, and unconscious (Reber, 1993; Perruchet, 2008)? Consistent with DeKeyser's (2007) transferability hypothesis, a pedagogical approach is proposed that converts learners' conscious declarative knowledge to unconscious procedural knowledge. In the model of learner progress above, learners are guided from an initial stage where competence is lacking and errors are made unconsciously to the stage of automaticity, where targets are produced intelligibly without the learner having to stop and think about it.

LEVEL 1:
Unconscious
incompetence

At this stage, you were probably making errors in listening and speaking without knowing what those mistakes were.

LEVEL 2:
Conscious
incompetence

As you learn, you may still make mistakes, but you're starting to know & understand what kinds of mistakes they are.

LEVEL 3:
Conscious
competence

At this level you may still need reminders or prompting from your teacher to help you think about your errors, but you know how to correct them.

LEVEL 4:
Unconscious
competence

This is your goal: At this level, you speak intelligibly and listen accurately. You do this spontaneously without needing to stop and think about it all the time.

Figure 1. The Four Levels of Competence: Learner's Perspective.

The proposed approach advocates intervention in the form of metalinguistic feedback at Level 3, the stage when learners can produce pronunciation (as well as syntactic and morpho-syntactic) targets on demand but have yet to integrate them into their spontaneous production. Two key elements are recommended to help solidify new speech patterns for learners: Teaching Talk & student Tell-Backs. Teaching Talk is defined as the succinct language of instruction used to introduce segmental/ suprasegmental concepts. It consists of learner-friendly chunks that teachers can use to help learners build their declarative knowledge. It is characterized as a minimalist statement of the rule or concept, presented before and again after explanations and examples, allowing it to be re-stated and retained by the learner. By way of analogy, like the PB&J or meat & cheese sandwiched between layers of bread, explanations and examples are sandwiched between layers of succinct, minimalist Teaching Talk. This example, introducing the components of stress for vowels in multi-syllabic words, uses a stress pattern notation system adapted from a 2004 study by Murphy & Kandil. (For additional Tips on syllables and a Checklist for Learning New Words, see Reed 2014.)

Teaching Talk Sample:

      Stressed syllables are longer, louder, higher, clearer.
           In most English words with more than one syllable, the syllables are not equal.
           When you learn a new word in English, you need to learn its stress pattern.
           That's because in English, assignment of stressed syllables is not predictable.
           For example, here are words for three musical instruments: piccolo, piano, violin.
           Each word has three syllables, but the stress patterns are different:
                  piccolo – 3 syllables with stress on the first: 3-1
                  piano – 3 syllables with stress on the second: 3-2
                  violin – 3 syllables with stress on the third: 3-3

So, when learning new words in English, learn which syllable is stressed.
Stressed syllables are longer, louder, higher, clearer.

Tell-Backs, a term borrowed from the literature on reading instruction (Vanderwood & Nam, 2007), constitute the language that learners use to re-state their understanding of a concept or pattern. Tell-backs may be verbatim, but are often reformulated, reflecting internalization of the concept. To illustrate the former, if you find yourself repeatedly recasting mispronounced –ed endings on regular past tense/participle verbs, only to hear your learners incorrectly add an extra syllable to the same or other verbs, consider using this Checklist (Reed & Michaud, 2005).

Checklist:
How do you say the "–ed" ending on regular past tense verbs?

Look: Find the simple (root) form of a verb, without any endings.
Ask: What is the final sound (not letter)?
Is it: /t/ or /d/?

    If *yes*. . . ✔
        Add an extra syllable.
            ⇒ Say "-ed" as [Id].
    If *no*. . .
        Ask: Is the final sound unvoiced?
          If *yes*. . . ✔
             There is no extra syllable.
             ⇒ Say "-ed" as [t].
          If *no*. . .
             There is no extra syllable.
             ⇒ Say "-ed" as [d].

So, when saying the past tense ending:

1. **Voiced** sounds use **voiced** endings, **[d]**.

2. **Unvoiced** sounds use **unvoiced** endings, **[t]**.

3. Sounds **/t/ or /d/** use an **extra syllable, [Id].**

Figure 2. Pronouncing –ed Endings on Regular Verbs.

Teaching Talk can take many forms. In addition to a short, clear definition of a key term or concept, it may be in the form of a question, as in "Is the final sound /t/ or /d/?" to remind learners of the –ed ending checklist. Teaching Talk is proposed to work best when it matches the language of corrective feedback (CF), conceptualized here as coaching learners to recall and retrieve what they know and put it into practice. That is, whatever metalinguistic feedback teachers offer ("Make the stressed syllable longer, louder, higher, clearer" or "No /t/ or /d/: No Extra Syllable" etc.) when prompting learners in the classroom should be the same language used to teach the concept or pattern to begin with. Teaching Talk has these advantages:

- consistency across class meetings throughout a semester of instruction
- transparency for learners
- increased metacognition for learners as they use the prompts to recall previously learned material, mentally run through checklists or strategies, and take responsibility for supplying the target form.

Teaching Talk is most efficient when it matches the language teachers elicit from learners in the form of student 'tell-backs,' which serve, in turn, to help leaners form new mental models and self-monitor for accuracy.

Table 2

Teachers' Companion to the Four Levels of Competence

Stages of Instruction	*Stages of Progress*	*Mechanisms of Progress*	*Measurement of Progress*
Beginning Instruction	Level 1: – – Students make errors unwittingly/unconsciously	Initial diagnostics; Teachers gather baseline data	
After initial instruction	Level 2: + – Students gain conceptual grasp; still make errors	Teaching Talk Guided Practice Principled CF Coaching	Student Tell-Backs
After Instruction & Practice	Level 3: + + Students master specifics of target sounds/structures	Guided (scaffolded) Practice Principled CF Coaching	Teacher- or Peer-Prompted Production Student Self-Correction
After Scaffolded Practice	Level 4: – + New mental models Automatized knowledge	CF Coaching Prompted Production	Spontaneous target-like production
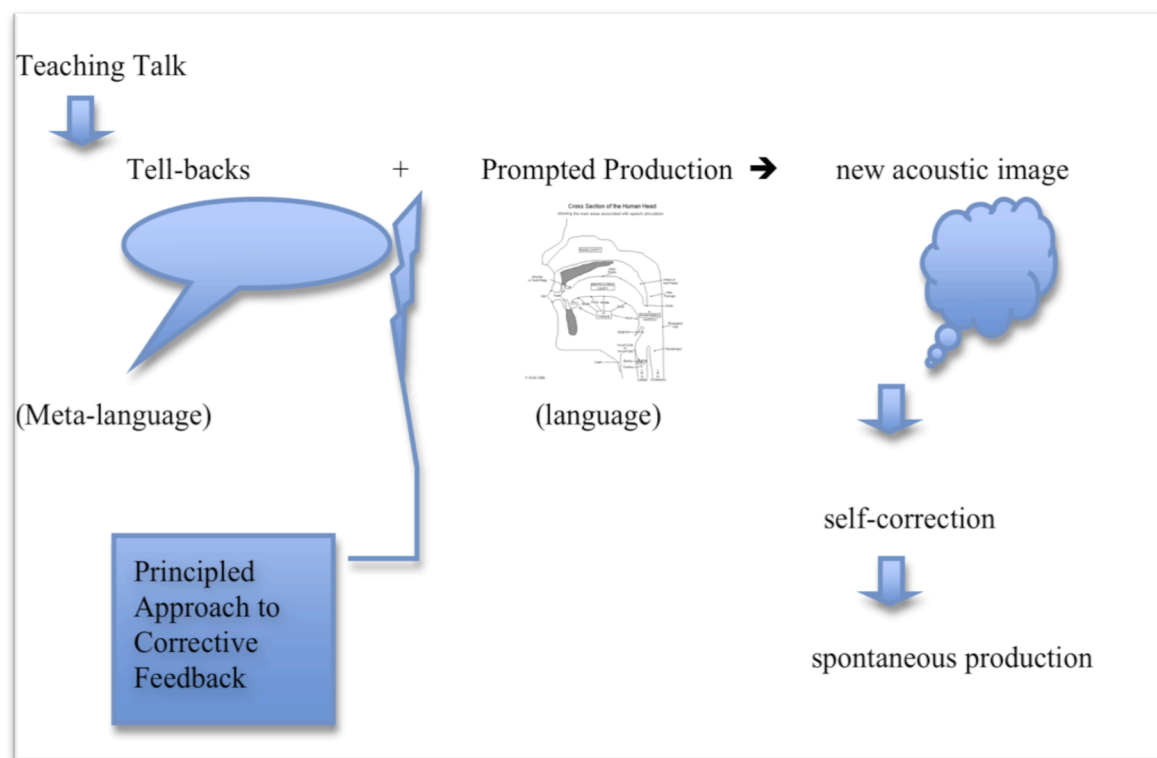
Figure 3. Teaching Talk, Tell Backs, and Prompted Production.

This metacognitive coaching process can be schematized in a flow chart (Reed & Michaud, 2010, p.35) highlighting the role of instructor-prompted production as a mediating interface between learners' declarative knowledge and their proceduralized target-like spontaneous production.

Whether, and if so how, learners can transfer their declarative knowledge (Level 2: what one knows consciously) to procedural knowledge (Level 4: what one can produce unconsciously) has long been a topic of debate. Contrasting views on explicit to implicit knowledge transfer can be found not only among scholars but even within the same model of acquisition. Krashen, for instance, took a non-interface position when proposing his acquisition versus learning distinction, claiming that learned knowledge can not be converted to acquired knowledge (1981; 1982, p. 83; 1985 pp. 42-3: "learning cannot turn into acquisition"). Yet the Monitor in his model (Krashen, 1985) allows for retrieval of knowledge learned in instructed settings under three conditions: time, focus on form, and rule knowledge. In his Transferability hypothesis, DeKeyser (2007) proposes a slightly overlapping set of conditions: time, meaningful practice, and sufficient input. In addition to ample input, output has also been suggested as facilitative. Larsen-Freeman (2003) points out that because of the synchronous nature of doing and learning, output practice does more than "simply serve to increase access to previously acquired knowledge" (p. 114). As noted by de Bot (1996), while output does not create completely new declarative knowledge, it "plays a direct role in enhancing fluency by turning declarative knowledge into procedural knowledge" (p. 553). Fluency, however, often consists of fossilized output, argued by Ellis (1989) to occur when learners have achieved communicative adequacy. To ensure that declarative knowledge is not by-passed during production, intervention is required to promote accuracy and intelligibility. Corrective feedback, conceptualized here as prompted production and delivered unobtrusively as pronunciation coaching, scaffolds practice and thus is empowering, rather than embarrassing for learners. Pronunciation coaching is achieved when the language of intervention matches the language of instruction used to introduce a target feature, and is in turn consistent with the tell-backs that learners use to internalize those features.

This Teaching Tip is intended to identify an interface between learners' declarative knowledge and their spontaneous procedural knowledge. The Teaching Tip proposes metalinguistic pronunciation coaching as an essential element in a metacognitive approach to bridge the gap between learners' explicit knowledge of a rule or feature of English and target-like spontaneous production. Three interface mechanisms are proposed to help learners achieve automaticity, or unconscious competence with the target materials:

> Teaching Talk: learner-friendly succinct form-focused Language of Instruction
> > • Establishes explicit, declarative knowledge
> Tell-Backs: learner-generated restatements of the minimalist teaching talk chunks
> > • Facilitates internalization of form-focused declarative knowledge
> Pronunciation Coaching: succinct, minimalist, learner-friendly Corrective Feedback
> > • Matches Teaching Talk to Tell-Backs to prompt self-monitoring, self-correction

CONCLUSION

In summary, this Teaching Tip promotes the use of unobtrusive corrective feedback in the form of pronunciation prompting that uses language that is uniform, delivered as Teaching Talk to establish declarative knowledge and restated by learners as Tell-Backs for the purpose of internalizing the concepts and converting explicit to procedural knowledge.

ABOUT THE AUTHOR

Marnie Reed is Associate Professor of Education and affiliated faculty in the Program in Applied Linguistics at Boston University. She is the director of the graduate program in Teaching English to Speakers of Other Languages (TESOL) in the School of Education. The focus of Dr. Reed's research, publications, and conference presentations is second language acquisition, specifically in applied phonetics and phonology. Her current area of interest and research is in the role of metacognition in cross-linguistic awareness of the pragmatic functions of intonation. She is co-author (with Christina Michaud) of *Sound Concepts: An Integrated Pronunciation Course* (McGraw-Hill), and co-editor (with John Levis) and chapter contributor to the Wiley-Blackwell *Handbook of English Pronunciation*.

REFERENCES

Bialystok, E. (1982). On the relationship between knowing and using linguistic forms, *Applied Linguistics 3*, 181–206.

de Bot, K. (1996) The psycholinguistics of the output hypothesis. *Language Learning, 46*, 529 - 555.

DeKeyser, R. M. (2003). Implicit and explicit learning. In C. J. Doughty & M. H. Long (Eds.). *The Handbook of Second Language Acquisition* (pp. 313-348), Oxford: Oxford University Press.

DeKeyser, R. M. (2005). What makes learning second-language grammar difficult? A review of issues. *Language Learning, 55*(Supplement 1), 1-25.

DeKeyser, R. M. (2007). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories In Second language Acquisition* (pp. 97-113). *NJ: Erlbaum*

Derwing, T.M., & Rossiter, M.J. (2002). ESL learners' perceptions of their pronunciation needs and strategies. *System, 30*, 155-166

Ellis, N.C. (2005) At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition 27*, 305–352.

Ellis, R. (1989). Are classroom and naturalistic acquisition the same? *Studies in Second Language Acquisition, 11*(3)*, 305-328*

Flege, J. E. (1993). Production and perception of a novel, second-language phonetic contrast. *Journal of the Acoustical Society of America, 93*, 1589-1608.

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

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.

Krashen, S.D. (1981) Second Language Acquisition and Second Language Learning Oxford: Pergamon Press.

Krashen, Stephen D. 1982. *Principles and practice in second language acquisition*. Oxford: Pergamon Press.

Krashen, S. (1985). *The input hypothesis: Issues and implications*. New York: Longman.

Larsen-Freeman, D. (2003). *Teaching language: From grammar to grammaring*. Boston, MA: Heinle.

Levis, J. (2015). Learners' views of social issues in pronunciation learning. *Journal of Academic Language & Learning, 9*(1) A42-A55.

Moyer, A. (2014). Exceptional outcomes in L2 phonology: The critical factors of learner engagement and self-regulation. *Applied Linguistics, 35*(4), 418-440.

Murphy, J. & Kandil, M. (2004). Word-level stress patterns in the Academic Word List. *System, 32*(1)*,* 61-74.

Perruchet, P. (2008). Implicit learning. In J. Byrne & H. L. Roediger, III (Eds.), *Learning and memory: A comprehensive reference* (Vol. 2, pp. 597–621). Oxford, England: Elsevier.

Reber, A.S. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious*. New York: Oxford University Press.

Reed, M. (2014). The English syllable: Big news, bad news and its importance for intelligibility. In J. Levis & S. McCrocklin (Eds). *Proceedings of the 5th Pronunciation in Second Language Learning and Teaching Conference* (pp. 189-202). Ames, IA: Iowa State University.

Reed, M., Michaud, C. (2005). *Sound concepts: An integrated pronunciation course.* NY: McGraw-Hill.

Reed, M. & Michaud, C. (2010). *Goal-driven lesson planning for teaching English to speakers of other languages.* Ann Arbor, MI: University of Michigan Press

Strange, W. (1995). Cross-language studies of speech perception: A historical review. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (pp. 3-45)*.* Baltimore, MD: York Press.

Vanderwood, M. L., & Nam, J. (2007). Response to intervention for English Language Learners: Current developments and future directions. In S. R. Jimerson, M. K. Burns & A. M. VanDerHeyden (Eds.), *The Handbook of Response to Intervention: The Science and Practice of Assessment and Intervention* (pp. 408-417). NY: Springer.

TEACHING TIPS

PROVIDING INDIVIDUALIZED HOMEWORK
AND ACCOUNTABILITY FOR ITAS VIA INTERNET RESOURCES

Monica Richards, Iowa State University

Pronunciation instructors of international teaching assistants (ITAs) frequently provide individual feedback highlighting the fluency, suprasegmental, and segmental challenges most likely to inhibit a particular ITA's successful interaction with undergraduates. Yet providing ITAs with practice actually implementing individual feedback given, adequate to enabling their development of new, more communicatively effective pronunciation habits, remains difficult. However, the Internet and learning management systems (LMSes, e.g., Moodle) contain resources capable of supporting and holding students accountable for focused, self-directed work on nearly any pronunciation-related target. This article briefly overviews the challenge faced by ITA instructors of advanced pronunciation, the feasibility of individualizing assignments in advanced pronunciation classes, the second language acquisition (SLA) foundations for a series of exercises designed to address common ITA difficulties, and sample directions for all exercises described.

INTRODUCTION

One challenge faced by teachers in advanced pronunciation courses for ITAs is that while their students' overall intelligibility may be greater than that of students in lower-level pronunciation courses, there is often less uniformity in what advanced students' specific areas of weakness are. For example, while some students in advanced pronunciation courses may need continued work on fluency, others may be *extremely* fluent but difficult to understand because their distribution of pause units (Brown, 1977, 2011), phrase stress, and/or intonation are different than North American English-speaking students expect and therefore can easily process. In addition, many times advanced pronunciation courses include a few students who are quite strong prosodically, but whose persistent difficulty in pronouncing a small set of problem segmentals renders them sometimes incomprehensible. Also, such courses may include students who struggle not so much with any particular segmental as with the tendency in English phonotactics to tack one consonant upon another at the end of syllables. These students' pronunciation, therefore, is not only difficult to understand, but also may imply to listeners that they are weaker in English grammar than they actually are, since their production of key grammar markers such as "-s" and "-ed" is spotty at best. Finally, probably every advanced pronunciation course includes students who have tested into the course not so much because of pronunciation problems, but because of weakness in listening or pervasive grammar issues.

The challenge faced by teachers of advanced pronunciation courses then is to provide their students with enough instruction and practice specific to their particular areas of weakness so students can make substantial progress during the relatively brief period of the course and, ideally, learn how they can continue to develop their weak areas even after the course has concluded, whenever they feel the

need (e.g., as students near graduation and want to strengthen their competitive edge on the North American job market!). However, how can this be accomplished when instruction and practice that barely overview what some students need end up being overkill for others? After all, students understandably tend to have a very limited tolerance for instruction and activities they view to be "wasting their time."

The obvious answer is providing students individualized instruction and practice, but since pronunciation instructors neither have time nor are paid to tutor their students, the bulk of the individualization that advanced pronunciation students actually receive tends to be limited to individual feedback on their fluency, suprasegmental, segmental, listening, grammatical, and pedagogical challenges revealed by the ITA test that placed them into the course; to individual feedback on in-class presentations and interactions; and to occasional individual mini-lessons dealing with particularly egregious errors or items about which a student specifically asks. That is, while it is possible many advanced pronunciation students receive adequate *information on what their individual areas of weakness are*, it is highly unlikely such students are being given *practice opportunities adequate to enabling their development of new, more communicatively effective pronunciation, listening, or grammar.* This paper therefore provides teachers with ideas based on several Web-available, ready-to-use resources for how they can provide (and hold students accountable for!) substantial practice of continuing weak areas within the sometimes severe constraints instructors face for how much time they can afford to invest in a given course.

How Can I Hold Students Accountable for Completing Individually Assigned Homework?

The major question instructors are likely to have in relation to the proposal that they provide advanced pronunciation students with individualized assignments is how to hold students accountable for *completing* their assignments. While there is probably no foolproof way of providing accountability, utilizing a resource like the Moodle learning management system's (LMS's) add-on "questionnaire," with its "respond daily" questionnaire type, allows teachers to provide students a reporting tool they can use only once per day. My questionnaire (Appendix 1) requires students to respond to two questions: 1) "Have you completed your 15-minutes-per-day homework today?" with the radio button options of "Yes" and "No" and 2) "Which assigned exercise or other allowable homework activity did you work on?" with an 80-character text input box in which students indicate the exercise they did. In order to pass, students must report having completed 15-minutes-per-day assignments at least 60 days throughout the semester.

Of course, students *can* lie one or more of the required 60 days, saying "yes" and listing some exercise they actually *haven't* worked on that day, but when I introduce the 15-minutes-per-day assignment at the beginning of the semester, I always remind my students that it is unreasonable for them to hope in the future to pass the ITA test that put them into my class (thereby avoiding being required to take future pronunciation classes) if they don't improve their pronunciation; that improving their pronunciation will probably also benefit their post-grad-school dream careers; and that the total pronunciation homework my course demands is relatively low (60 days of 15-minutes-per-day homework, which adds up to less than an hour per week, plus whatever time they need to prepare their various course presentations). In addition, each time I meet with students throughout the semester I check how many 15-minutes-per-day reports they have submitted and what they say they have been doing relative to their individualized assignments. In our meetings, we also talk about whether or not they are seeing progress through their 15-minutes-per-day exercises and if not, what

changes we can make to their assignments so they *do* see progress. By my second semester of having individualized 15-minutes-per-day homework as a major component of my course, I became convinced that most, if not all, of my students *were* actually doing the homework they reported each day. While in my courses I have utilized an LMS-based accountability tool, accountability could also be managed in other ways, e.g. via an email rule automatically sorting all emails from students with the subject line beginning "Today's 15 minutes completed." The remainder of the subject line could be the name of whichever exercise the student did.

Overviewing a Series of Exercises Addressing Common ITA Difficulties

At the beginning of the semester, I provide students individual feedback via Moodle's forum activity (in "separate group" mode" with one student per group; feedback could also be given via email). My feedback not only categorizes and prioritizes each student's particular strengths and weaknesses as identified by their ITA test, but also includes for each problem category (e.g., fluency, phrase stress, intonation, etc.) links to relevant 15-minutes-per-day exercises targeting that specific problem category. Several of my most commonly recommended 15-minutes-per-day exercises, the directions for which are presented to students via static pages on our course website, and their grounding in the primarily interactionist and cognitive second language acquisition (SLA) literature, are overviewed below. These exercises are designed to maximize student motivation (Dickinson, 1987) by allowing learners to use as input any of a wide variety of high-interest source materials such as TED Talks (ted.com/talks/) and Newsy reports (newsy.com). Most of my recommended exercises reflect Gass' (1997) integrated model of second language acquisition, in which learners 1) realize there is a gap in their language knowledge (the "apperception stage"), 2) comprehend input (the "comprehended input" stage), 3) compare their interlanguage to the input (the "intake" stage), 4) move to integrate the intaken feature into their L2 interlanguage (the "integration" stage), and 5) produce output demonstrating acquisition (the "output" stage).

The first exercise, *4/3/2* (Appendix 2), is well-respected in the literature as a fluency-development exercise (Maurice, 1983; Nation, 1989) because it grows students' "online planning" capacity (R. Ellis, 2005) by its demand for "pushed output" (Skehan, 1998b, Swain & Lapkin, 1995) and in addition, develops their ability to reduce the hyperactive language monitoring (Krashen, 1981; Levelt, 1983; Morrison & Low, 1983) that often generates an overabundance of repairs and other disfluencies. *4/3/2* also develops in learners increasingly automatized language and discourse knowledge by its demand for repeated production on a single topic (Skehan, 1998a). My 15-minutes-per-day adaptation of *4/3/2*, however, under the pressure of rendering it realistic for regular self-tutoring use, will undoubtedly disgruntle teachers having strongly sociocultural SLA bents, since it makes optional the authentic multiple-interlocutor communication built into Maurice's (1983) original conception of the activity.

The second exercise, *Fluency Buildup* (Appendix 3), reduces the challenge of fluency development to an absolute minimum, namely pause-unit-sized chunks (Brown, 1977, 2011), and can scaffold (Wood, Bruner & Ross, 1976) fluency development for students with stabilized/fossilized (Long, 2003) habits of hyperactive language monitoring (Krashen, 1981; Levelt, 1983; Morrison & Low, 1983) as well as phrase stress and intonation development for students whose ingrained phrase stress or intonation patterns render their spoken English difficult for their North American target listeners to understand. *Fluency Buildup* is also valuable because of the face validity it has for learners whose educational cultures place a high value on imitation as a means of learning.

The third 15-minutes-per-day exercise, *Shadowing* (Appendix 4)*,* also involves following a model, but allows learners to use expert speaker talk directly as a scaffold (Wood, Bruner & Ross, 1976) toward producing more comprehensible output (Swain, 1995) while pushing learners to match the

talking speed of the speaker they are shadowing, *Shadowing* thus increases fluency and develops learners' capacity to subdue an overactive language monitor (Krashen, 1981; Levelt, 1983; Morrison & Low, 1983) as well as supports their development of more target-like phrase stress and intonation.

The fourth exercise, *Analyze2Imitate: Pause Units, Phrase Stress, and Intonation* (Appendix 5), requires students to be far more analytical than any of the previous exercises and aims to disrupt stabilized (Long, 2003) pause unit (Brown, 1977, 2011), phrase stress and intonation patterns problematic for a student's target listeners by requiring learners to develop via a transcript of a speaker's talk their own "enhanced input" (Sharwood Smith, 1993) for use in imitating that talk. *Analyze2Imitate*'s form-focused (R. Ellis, 2002; Long, 1991; Long & Robinson, 1998) input enhancement process requires learners to notice (Schmidt, 2001) and mark their chosen speaker's 1) pause units, 2) phrase stress (highlighting all instances of default vs. non-default phrase stress in different colors), and 3) uses of non-default (i.e., non-falling) intonation (via underlining). This process obviously necessitates that learners engage in repeated listening which, according to the input frequency research (N. Ellis, 2002), should also contribute to learners' noticing of target forms (Schmidt, 2001). *Analyze2Imitate* then requires learners to use their enhanced transcript to guide their recorded re-enactment of the speaker's talk and then assess their recording against their enhanced transcript to identify where they successfully imitated the expert speaker and where they did not. I have found that learners appreciate the immediately visible learning brought about by this somewhat complex exercise.

The final exercise introduced in this paper, *Analyze2Imitate: Ending Consonants* (Appendix 6) basically reflects the same SLA theoretical and research foundation as *Analyze2Imitate: Pause Units, Phrase Stress, and Intonation* (Appendix 5) except that the *Ending Consonants* exercise may not require students to engage in repeated listenings because its input enhancement (Sharwood Smith, 1993) demands are relatively simple, namely highlighting all consonants following the last pronounced vowel for all words in a transcript. The aim of the *Ending Consonants* exercise is to help learners acquire the ability demanded by both English vocabulary and English grammar of appending one consonant after another at the end of words, a pronunciation feature many English language learners find difficult to acquire since the English tendency to close syllables with consonants is marked among the world's languages (Eckman, 1977, 2004). (Incidentally, the *Analyze2Imitate* design is useful for non-pronunciation purposes as well: Another 15-minutes-per-day exercise I recommend is *Analyze2Imitate: Grammar* (Appendix 7). This exercise has students highlight grammar features with which they tend to make mistakes, e.g. highlighting all verbs if they have trouble with verb tenses or subject/verb agreement or all instances of the article "a" and plural "-s" if they have trouble using the countable noun markers that English requires.)

Obviously, underlying all *Analyze2Imitate* exercises are the emergentist (cf., N. Ellis, 2002, 2007) and skill-acquisition (DeKeyser, 1998; Johnson, 1988, 1996) theories' claims that conscious, explicit, declarative language knowledge *can* become unconscious, implicit, and procedural and that intentional learning *can* lead to acquisition, not merely "learning" (Krashen, 1994). Other emergentist assumptions (cf., N. Ellis, 2002, 2007) instantiated in the *Analyze2Imitate* exercises are that 1) acquisition involves gradual strengthening of associations, e.g., associations between a speaker's meaning and his or her use of contrastive/emphatic stress and non-default intonation, and 2) language in the brain is primarily rooted in examplars, not rules (though learners <u>do</u> inductively derive rules from exemplar patterns).

CONCLUSION

The aim of this paper has been to demonstrate to instructors of advanced pronunciation students that they do *not* need infinite amounts of motivation, time, energy or any other resource to provide students such as ITAs with substantial SLA-theory-and-research-grounded practice that wholly parallels their widely varying individual needs. In our era of increasing selection in high-quality Internet resources that are easily turned to language-learning advantage, individual student feedback linked to appropriate SLA-grounded exercises will suffice if combined with an appropriate accountability mechanism. Such exercises will not only build students' language capacity during our courses, but provide them learner training for continued independent language acquisition. So let's individualize our ITA students' homework!

ABOUT THE AUTHOR

Monica Richards is a Ph.D. candidate in Applied Linguistics and Technology at Iowa State University. Her research and materials development interests relate to SLA-grounded pedagogy for all skills, particularly pronunciation. All exercises below — in addition to other recommended 15-minutes-per-day exercises on pronunciation, grammar, listening, etc. — are available for linking on her website relateworldwide.org.

REFERENCES

Brown, G. (1977). *Listening to Spoken English*. London: Longman.

Brown, G. (2011). Prosodic structure and the given/new distinction. In: Cutler, A., Ladd, D. R. (Eds.), *Prosody: Models and Measurements*. Berlin: Springer-Verlag, pp. 67-77.

DeKeyser, R. (1998). Beyond focus on form: Cognitive perspectives on learning and practicing second language grammar. In C. Doughty, & J. Williams (Eds.), *Focus on Form in Classroom Second Language Acquisition* (pp. 42–63). Cambridge, UK: Cambridge University Press.

Dickinson, L. (1987). *Self-Instruction in Language Learning*. Cambridge: Cambridge University Press.

Eckman, F. (1977). Markedness and the contrastive analysis hypothesis. *Language Learning, 27*, 315-330.

Eckman, F. (2004). Universals, innateness and explanation in second language acquisition. *Studies in Language Acquisition, 28*, 682–703.

Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition, 24,* 143-188.

Ellis, N. C. (2007). The weak-interface, consciousness, and form-focused instruction: Mind the doors. In S. Fotos & H. Nassaji (Eds.), *Form-Focused Instruction and Teacher Education: Studies in Honour of Rod Ellis*, pp. 17-33, Oxford: Oxford University Press.

Ellis, R. (2002). Does form-focused instruction affect the acquisition of implicit knowledge? *Studies in Second Language Acquisition, 24*(2), 223-236.

Ellis, R. (2005). *Planning and Task-Performance in a Second Language*. Amsterdam: John Benjamins.

Gass, S. M. (1997). *Input, Interaction, and the Second Language Learner*. Mahwah, NJ: Lawrence Erlbaum.

Johnson, K. (1988). Mistake correction. *ELT Journal, 42*(2), 89–96.

Johnson, K. (1996). *Language Teaching and Skill Learning*. Oxford: Blackwell.

Krashen, S. (1981). *Second Language Acquisition and Second Language Learning*. Oxford: Pergamon Press.

Krashen, S. (1994). The Input Hypothesis and its rivals. In C.N. Ellis (Ed.), *Implicit and Explicit Learning of Languages* (pp.45-77).

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition, 14*, 41-104.

Long, M. (1991). Focus on form: A design feature in language teaching methodology. In: K. de Bot, R. Ginsberg, & C. Kramsch (Eds.), *Foreign language research in cross-cultural perspective* (pp. 39-52). Amsterdam: John Benjamins.

Long, M. H., & Robinson, P. (1998). Focus on form: Theory, research and practice. In Doughty, C. J., & Williams, J. (eds.), *Focus on Form in Second Language Acquisition* (pp. 15-41). Cambridge: Cambridge University Press. Reprinted in Ortega, L. (ed.), *Second Language Acquisition: Critical Concepts in Linguistics*. London: Routledge.

Long, M. H. (2003). Stabilization and fossilization in interlanguage development. In Doughty, C. J., & Long, M. H. (eds.), *Handbook of Second Language Acquisition* (pp. 487-535). Oxford: Blackwell.

Maurice, K. (1983). The fluency workshop. *TESOL Newsletter, 17*(4), 29.

Morrison, D. and Low, G. (1983). Monitoring and the second language learner. In: Richards, J. and Schmidt, R. (eds.), *Language and Communication* (pp. 228-250). London: Longman.

Nation, P. (1989). Improving speaking fluency. *System, 17*, 377-384.

Nation, P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 3–32). New York: Cambridge University Press.

Sharwood Smith, M. A. (1993). Input enhancement in instructed SLA. *Studies in Second Language Acquisition, 15*, 165–179.

Skehan, P. (1998a). *A Cognitive Approach to Language Learning*. Oxford: Oxford University Press.

Skehan, P. (1998b). Task-based instruction. *Annual Review of Applied Linguistics, 18*, 268-286.

Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principles and Practice in Applied Linguistics* (pp. 125–144). Oxford: Oxford University Press.

Swain, M. & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step toward language learning. *Applied Linguistics, 16*, 371-391.

APPENDIX 1
Moodle's "Questionnaire" activity used as a daily homework reporting tool ([Return to text](#))

APPENDIX 2
Directions for sample 15-minutes-per-day homework assignment *4/3/2*
(adapted from Paul Nation's *Learning Vocabulary in Another Language*, 2001) (Return to text)

4/3/2 directions

Directions for building fluency via 4/3/2

1. Choose one of the following to teach/present:

- Choose a concept/problem/example in your field that you could explain to an undergrad in just a few minutes
- Choose 2-3 slides from one of your class or conference presentations
- Choose 2-3 slides from an interesting presentation you find on Slideshare, Slideboom, or Speaker Deck or a few frames from an interesting presentation you find at http://prezi.com/explore/
- Choose a brief online article discussing interesting information you might want to tell someone in the future

2. Unless you're using existing slides or Prezi presentation frames, spend 5-15 minutes <u>outlining</u> **the main points/important details you want to mention.** (Do not write a transcript!!!)

3. Set the countdown timer below for 4 minutes and give your talk (either to a partner or to an imaginary audience. If you give your talk to an imaginary audience, <u>don't</u> let yourself start over if you make a mistake! Just continue on—you can correct your mistake when you give your talk in 3 minutes.)

4. Set the timer for 3 minutes and give your talk again.

5. Set the timer for 2 minutes and give your talk again. (Can you sense that this time you were more fluent and more accurate—and you used more complex language—than when you gave your talk the first time?)

www.online-stopwatch.com

Stop Watch   Count Down

(FYI: It's great if you can do this exercise with three other people, because then you can rotate partners once you have both given a talk to and heard a talk from your current partner. Doing this exercise in a group of 4 allows all of you to give your talk to a new "real audience" every time the amount of time for your talk changes.)

APPENDIX 3

Directions for sample 15-minutes-per-day homework assignment *Fluency Buildup* ([Return to text](#))

Fluency Buildup

If your OECT Feedback indicates you struggle with fluency,

1. **Find a TED Talk (e.g. from Monica's list of recommended 3-minute TED Talks) or Newsy report** you find interesting.
2. Select the transcript text with your mouse and click "Ctrl + C" to copy it. Paste your copied text into a Microsoft Word or other document by clicking Ctrl + V.
3. **Watch the TED talk at least twice, inserting "/" where the speaker makes a short pause and inserting "//" where the speaker makes a long pause.** (Use pencil if you do this with a printed transcript, so you can easily fix mistakes!)

> When you see / a diagram like this, / I don't want you to be afraid. I want you to be excited. I want you to be relieved. / Because / simple answers may emerge. / We're discovering in nature that / simplicity often lies / on the other side of complexity. So for any problem, / the more you can zoom out / and embrace complexity, the better chance you have / of zooming in / on the simple details / that matter most. (Eric Berlow: "Simplifying Complexity," TED)

4. **Listen again to how the speaker says a sentence or two and then pause the video.**
5. **Say the first phrase 5 times, without any pauses and following the speaker's phrase stress, intonation, etc.** (If you do pause when saying the phrase, start over at 1 again when counting your 5 times.), e.g.

> What you say: "When. . . .you see. . . ."
> *What you think: 'Oh, no! I paused. I need to start over. . .'*
> What you say: "When you see. . . , When you see. . ., When you see. . ., When you see. . ., When you see. . . ."
> *What you think: 'Okay, I can go to the next phrase. . . .'*

6. **Do the same thing with the second phrase 5 times.**
7. **Now say the first and second phrase together 3 times, pausing only between the phrases:**

> What you say: "When you see / a diagram like this,. . . , When you see / a diagram like this,. . . , When you see / a diagram like this,. . . , "

8. **Say the third phrase 5 times** as above.
9. **Now say the first, second, and third phrase together 3 times.**
10. **Continuing adding one phrase at a time,** listening to the video when needed a few sentences at a time, **until you can say an entire paragraph fluently, pausing only where the speaker paused.**
11. **Start over with the next paragraph at step 6 above. Once you've finished practicing the second paragraph, practice saying both the first and second paragraphs 3 times,** allowing yourself to pause only where the speaker did.
12. **Continue this process until you can give the speaker's entire speech fluently.**

APPENDIX 4

Directions for sample 15-minutes-per-day homework assignment *Shadowing* ([Return to text](#))

Shadowing

If your OECT Feedback indicates you should work on your English pause units, phrase stress, and/or intonation via shadowing, watch the video below to learn about this technique:

Then, find a **TED Talk** (e.g. from **Monica's list of recommended 3-minute TED Talks**) or a **Newsy** report you find interesting and shadow the speaker's pause units, phrase stress (i.e. lengthening of the stressed vowel in the final key word of the pause unit), intonation (i.e. the rise and fall of his or her voice), etc.

APPENDIX 5
Directions for sample 15-minutes-per-day homework assignment
Analyze2Imitate: Pause Units, Phrase Stress, and Intonation ([Return to text](#))

Analyze2Imitate: Pause Units, Phrase Stress, and Intonation

1. **Find a TED Talk (e.g. from Monica's list of recommended 3-minute TED Talks) or a Newsy report** you find interesting.
2. Select the transcript text with your mouse and click "Ctrl + C" to copy it. Paste your copied text into a Microsoft Word or other document by clicking Ctrl + V.
3. **Watch the TED talk at least twice, inserting "/" where the speaker makes a short pause and inserting "//" where the speaker makes a long pause.**
4. Highlight **the last "important word" of each pause unit (usually a noun, verb, adjective, or adverb),** since this is the default position for phrase stress in English.
5. **Listen to the talk one more time to check whether the speaker** actually **stresses in each pause unit the word(s) you highlighted. Change your highlighting as necessary to match the speaker's** actual **phrase stress.** (FYI: English speakers tend to add phrase stress to a word by making whichever of its vowels receives the word stress longer.)
6. **If the speaker stresses a word in order to contrast it with another word,** highlight in a different color **the specific** syllable **he/she actually stresses** (It's important to highlight the exact syllable the speaker stresses and not just the word since contrast stress can not only change the word that gets stressed, but also which syllable in that word gets stressed). For example, Eric Berlow in his TED talk, "Simplifying Complexity," consistently shifts his word stress for the word "complicated" to contrast it with the word "complex":

   "So / I hope to convince you that complex / doesn't always equal complicated. // So for me /, a well-crafted baguette, fresh out of the oven /, is complex /, but a / curry / onion / green olive / poppy cheese bread is complicated" (Eric Berlow: "Simplifying Complexity," TED).

7. **Read the transcript again,** underlining **each pause unit you think the speaker will mark with a rising tone** vs. the default falling tone or other tones sometimes used in English.
8. **Listen again to the talk to check/correct your predictions.**

   "I also figured out / that / if you really want something badly enough, / you can do anything / for 30 days. // Have you ever wanted to write a novel? // Every November, / tens of thousands of people / try to write their own fifty thousand word novel / from scratch / in 30 days. // It turns out, / all you have to do / is write sixteen hundred and sixty-seven words a day / for a month. // So I did. // By the way, the secret / is not to go to sleep / until you've written your words for the day. // You might be sleep-deprived, / but / you'll finish your novel. // Now / is my book / the next great American novel? // No. / I wrote it in a month. / It's awful. / But / for the rest of my life, / if I meet John Hodgman at a TED party, / I don't have to say, / "I'm a / computer scientist." / No, no, / if I want to, I can say, / 'I'm a novelist.' /" (Matt Cutts: "Try something new for 30 days," TED).

9. **Record yourself giving the talk** (perhaps using a tool like http://online-voice-recorder.com/). Pay attention to pausing only where you marked either "/" or "//", to lengthening the stressed vowels in each of your highlighted words, and to using rising intonation with every pause unit you underlined.
10. **Check your recording.** Did you pause only where you marked "/" or "//"? Did you lengthen the stressed vowels of all your highlighted words? Did you use rising intonation for all of your underlined pause units? Mark in **bold** any pause units in which you made a pausing, phrase stress, or intonation mistake.
11. **Rerecord the talk**, paying particular attention to the stress and intonation of the words and phrases you marked in **bold**.

APPENDIX 6
Directions for sample 15-minutes-per-day homework assignment
Analyze2Imitate: Ending Consonants ([Return to text](#))

Analyze2Imitate: Ending Consonants

If your OECT Feedback indicates you frequently omit final consonant sounds:

1. **Find a TED Talk (e.g. from Monica's list of recommended 3-minute TED Talks) or a Newsy report** you find interesting.
2. Select the transcript text with your mouse and click "Ctrl + C" to copy it. Paste your copied text into a Microsoft Word or other document by clicking Ctrl + V.
3. Watch the TED talk at least twice, inserting "/" where the speaker makes a short pause and inserting "//" where the speaker makes a long pause. (Use pencil if you do this with a printed transcript, so you can easily fix mistakes!)
4. Read through the talk, highlighting all final consonant sounds like in the example below:

> When you see / a diagram like this, / I don't want you to be afraid. I want you to be excited. I want you to be relieved. / Because / simple answers may emerge. / We're discovering in nature that / simplicity often lies / on the other side of complexity. So for any problem, / the more you can zoom out / and embrace complexity, the better chance you have / of zooming in / on the simple details / that matter most. (Eric Berlow: Simplifying Complexity, TED)

APPENDIX 7
Directions for sample 15-minutes-per-day homework assignment
Analyze2Imitate: Grammar ([Return to text](#))

Analyze2Imitate: Grammar

The first step in improving your spoken English grammar is to increase your ability to hear errors <u>as</u> errors. One way you can develop this ability is the following:

1. **Find a TED Talk (e.g. from Monica's list of recommended 3-minute TED Talks) or a Newsy report** you find interesting.
2. Select the transcript text with your mouse and click "Ctrl + C" to copy it. Paste your copied text into a Microsoft Word or other document by clicking Ctrl + V.
3. **Record yourself reading the transcript aloud** (perhaps using a tool like http://online-voice-recorder.com/), **as naturally as possible, pretending you really <u>are</u> giving the talk.**
4. **Highlight in different colors all potential locations of the error types you know you have trouble with** (e.g. Highlight all verbs in your transcript in green if you know you have trouble with verb tenses and/or subject/verb agreement. Highlight all plural nouns in your transcript in yellow if you tend to say things like "two factor_." Highlight all articles—"a" and "an" and "the"— in orange if you know you tend to make article mistakes.)

> "Unless we do something to prevent it, over the next 40 years we're facing an epidemic of neurologic diseases on a global scale. A cheery thought. On this map, every country that's colored blue has more than 20 percent of its population over the age of 65. This is the world we live in. And this is the world your children will live in. For 12,000 years, the distribution of ages in the human population has looked like a pyramid, with the oldest on top. It's already flattening out. By 2050, it's going to be a column and will start to invert. This is why it's happening. The average lifespan's more than doubled since 1840, and it's increasing currently at the rate of about five hours every day. And this is why that's not entirely a good thing: because over the age of 65, your risk of getting Alzheimer's or Parkinson's disease will increase exponentially. By 2050, there'll be about 32 million people in the United States over the age of 80, and unless we do something about it, half of them will have Alzheimer's disease and three million more will have Parkinson's disease." (Gregory Petsko: "The coming neurological epidemic," TED).

5. **For each highlighter color you used, listen carefully to your recording one time, paying close attention to what you <u>actually</u> said for the words highlighted in that color vs. what the transcript indicates you <u>ought</u> to have said. Mark in bold any highlighted words on your transcript with which you made an error.**
6. **Rerecord yourself giving the talk one more time, paying particular attention to any words marked in bold on your transcript.** (But you don't need to listen to/check this second recording again unless you want to.)

Caromawati, C. (2016). Corpus of Misunderstandings from the Asian Corpus of English [Review]. 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. 258-261). Ames, IA: Iowa State University.

WEBSITE REVIEW

Corpus of Misunderstandings from the Asian Corpus of English
Corry Caromawati, Iowa State University

INTRODUCTION

English as a Lingua Franca (ELF) as a field of study has been an interest of experts such as Seidlhofer, Jenkins, Firth, House, and Kirkpatrick to investigate further the features of Englishes spoken by the outer and expanding circles approximately since the mid-1990s. The implication of their studies are crucial to the language learning in those circles, and are beneficial for ELF users. The spoken language features investigated in that particular field ranges from the segmental to suprasegmental level. The main purpose of the investigation of these features is to help learners in that particular area be understood by the native speakers of English and other ELF users (intelligibility). When this purpose is not achieved, there most likely will be a misunderstanding between the participants which causes a communication breakdown. This specific subject in the field of ELF is the focus of *The Corpus of Misunderstandings from the Asian Corpus of English* website.

This website was created based on some studies of the creators and their project groups on misunderstanding occurrences when EFL users communicated. The analysis conducted was on the word and phrase levels. The main purposes of this website are to make the studies accessible to more people who are interested in the field and to provide an opportunity to listen to the data used in the studies (D. Deterding, personal communication, November 12, 2015). The data analysis is expected to be a significant contribution to an understanding of what causes misunderstandings in that particular context.

THE CREATORS OF THE WEBSITE

The Corpus of Misunderstandings from the Asian Corpus of English website is managed by David Deterding and Salbrina Sharbawi. Deterding is a professor in the faculty of arts and social sciences in the University of Brunei Darussalam (UBD). He has contributed a lot in the field of World Englishes, particularly Asian Englishes such as Singapore, Chinese, and Brunei Englishes. His studies have been published in the forms of books, journal papers, and websites. His co-worker in developing this particular website, Sharbawi, is a senior lecturer in the same faculty of the same university. She has been working in the area between Sociolinguistics and Phonetics, known as Sociophonetics. Her interest in variations in World Englishes has been shown by her contribution in the field including her research findings which revealed the fact

258

PSLLT Reviews

that unlike Singapore English, Brunei English is less uniform. These two key people are parts of a bigger team which collects corpus for the Asian Corpus of English (ACE) data to provide examples of misunderstandings in the communication between EFL speakers in Asia.

THE CONTENT OF THE CORPUS

There are two main data sets in this website: The Brunei data and the ASEAN data. These two sets of data are parts of the corpus collected for the Asian Corpus of English. The Brunei data is gathered from group discussions involving at least three people and interviews between a Bruneian female and various other speakers from Asia. This data is being collected by the team of seven members, which is based at the UBD led by Sharbawi. The other set is the ASEAN data collected by prof. Andy Kirkpatrick in 2004 and 2005 located at the SEAMEO Regional Language Centre (RELC) in Singapore. The data gathered from six groups of three or four from different background discussing about their experiences and impressions about Singapore. The duration of the discussion was 20 minute per session. The result of the study of misunderstanding conducted by Kirkpatrick and Deterding (2006) was published in the *World Englishes* journal entitled "Emerging South-East Asian Englishes and intelligibility". Another analysis focused on the communicative strategies adopted by the speakers in this data was performed by Kirkpatrick (2011) and published as a book chapter with the title "English as a Lingua Franca in ASEAN: A Multilingual Model".

THE WEBSITE CONTENT AND LAYOUT

Figure 1 is a screenshot from the website to provide illustration of how the website is organized. As shown, the heading of the website is "Corpus of Misunderstandings from the Asian Corpus of English" representing its content described previously in this paper. It has three menus on the top bar: Home, Transcription, and Contact Us. *Home* mainly describes the general purposes of the website including the information about the book in which the main study is elaborated, and links to the Asian Corpus of English main website. *Transcription* provides information about the transcription convention used both in the study and in the website. *Contact us* provides links to get more information about the key people (Deterding & Sharbawi) as well as their contact information. Frankly speaking, they were very responsive when contacted.

The sidebar consists of the data sets, Brunei and ASEAN, followed by their subcorpora. In the Brunei data set, there are nine pairs of interviews: China & Brunei, Hong Kong & Malaysia, Hong Kong & Taiwan, Indonesia & Malaysia, Indonesia & Taiwan, Japan & Brunei, Laos & Brunei, Nigeria & Brunei, and Taiwan & Malaysia. In each pair, there is a description of the participants' background (nationality and age), the context of how the data were gathered, and some acknowledgement. The list of misunderstanding examples within the session is provided on the left side. Navigating under each sample, the user of this website will get information about the nationality and age of the participants involved in the discussion and the context when the misunderstanding occurred. Additionally, the transcription and the audio of the conversation are also provided, enabling users of the website to read and listen to the data. It is important to

259

highlight that under the transcribed conversation, there is a set of explanation on what actually caused misunderstanding (as shown in Figure 1). I think this is the most important part of this website as users can learn what was going on in the excerpt of negotiation of meaning.

The same organization is applied under the ASEAN data set. However, since the corpus was collected from group discussion, the menu of the subcorpora is named ASEAN Gp 1 to Gp 6 (Gp = group) instead of the participants' nationalities. The description of the location and time of the data collection as well as the participants' gender and nationality are also provided under each group.

Figure 1. The layout and sample content of the website.

EVALUATION

In this website, it is emphasized that "most of the conversations proceeded smoothly, without frequent breakdowns in communication. While we focus on misunderstandings here, we should remember that these were the exception, not the norm, in these interactions between speakers from the various ASEAN countries." (Deterding & Sharbawi, 2015). It indicates that the use of 'misunderstanding' goes broader that what it is expected. As stated by Björkman (2015), "Deterding adopts a rather broad definition of 'misunderstanding', which includes not only situations where something the speaker says is misinterpreted but also where something the speaker says is not understood by the listener." In addition to that, the studies did not only focus on when the 'misunderstanding' occurred, but also how it was dealt and avoided (Deterding, 2013). This process of analysis and the term used might potentially bring into different

260

expectation and interpretation of the users when using the website. I am aware that the field of misunderstanding has been recognized in this field, but I think there should be an evaluation on whether the term matches what is analyzed and presented in this website.

Another thing to look at is the number of the participants involved in the studies and included in this website. It has a very little chance for the data to be claimed as representation of the Englishes of the participants' nationalities, as "it may have idiosyncratic patterns that may not occur in individual English varieties." (Deterding, 2013, p. 186). As a consequence, more participants to represent each country are needed. I believe that the use of technology would be beneficial to gather more data for this website improvement. As a cross-nations project, it is important to note that it needs a lot of parties and studies involved to enable the Englishes in this website more generalizable. Unfortunately, there are not recognized studies in this particular topic (D. Deterding, personal communication, November 12, 2015).

The last thing I would like to discuss about this website is in terms of its usefulness. As stated previously, the main purpose of this website is to give greater access to the data of some studies in South East Asian Englishes: Deterding (2013), Kirkpatrick (2011), and Kirkpatrick & Deterding (2006), indicating that its role is limited to be as a complement to those studies. Additionally, there is no indication that users can navigate and interact directly with the data to have concordances or conduct some newer corpus study. As a consequence, this website has very limited circle of users.

Regardless of the limitation of the website that I mentioned earlier, I am honored to be given a chance to evaluate this website. It gives me an opportunity to learn about studies conducted to investigate characteristics of the Englishses in South East Asia, including my home country, Indonesia. These studies are beneficial for researchers who are interested in this field so that they can contribute to the English language instruction in South East Asia. Their contribution will give better understanding to the English teachers in the region to help their students encounter potential constrains in learning English.

REFERENCES

Björkman, B. (2015). Deterding, David: Misunderstandings in English as a Lingua Franca. An analysis of ELF interactions in South-East Asia. *Journal of English as a Lingua Franca*, *4*(2), 385-389.

Deterding, D. (2013). *Misunderstandings in English as a lingua franca: an analysis of ELF interactions in South-East Asia* (Vol. 1). Walter de Gruyter.

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

Deterding, D. & Sharbawi, S. (2015. Corpus of Misunderstandings from the Asian Corpus of English. Retrieved from http://fass.ubd.edu.bn/research/CMACE/home/index.html

Kirkpatrick, A. (2011). *English as a Lingua Franca in ASEAN: A Multilingual Model*, Hong Kong: Hong Kong University Press.

Choi, Y. (2016). The handbook of English pronunciation [Review]. 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. 262-265). Ames, IA: Iowa State University.

REVIEW

Reed, M., & Levis, J. (2015). *The handbook of English pronunciation*. Oxford, Ox: John Wiley & Sons.

[YunDeok Choi](#), Iowa State University

The Handbook of English Pronunciation, edited by Marnie Reed and John M. Levis, is a comprehensive reference book that addresses English pronunciation from various perspectives. I selected this book for my book review with the hope to acquire essential knowledge on English pronunciation, especially on teaching English pronunciation, by reading a recently published book written in layman's terms, which purportedly does not require any professional background knowledge on the part of the reader to fully benefit from it.

This book consists of 28 chapters and, in these chapters, English pronunciation is canvassed in depth on six encompassing themes: history, description, discourse, varieties, acquisition and teaching, targeting general audience including both experts in language education fields: applied linguists and language teachers and non-experts. However, this review will focus only on eight chapters included in the Pronunciation Teaching part, which I have been most interested in.

The last section opens with a chapter written by established scholars in the field of English pronunciation, Murray Munro and Tracey Derwing. In this opening chapter, Munro and Derwing overview intelligibility in terms of how it has been defined in the field of English pronunciation, assessed through a variety of tasks and investigated in research. In addition, the issues about effective implementation of intelligibility-oriented pronunciation instruction in class are also discussed.

In the next chapter, Beth Zielinski pinpoints the false assumption underlying a segmental versus supra-segmental debate, which is prevailed in the field of English pronunciation. She strongly argues that the two pronunciation features are inter-connected constituents of an integrated prosodic system and thus intelligibility and comprehensibility cannot be determined if the effects of two entities are considered dichotomously. Personally, as an applied linguist, this chapter is eye-opening to me since I have never deeply thought about how the two features are connected and how the interconnection influenced intelligibility and comprehensibility. On the other hand, as a language teacher, one thing left much to be desired to me is no concrete suggestion for implementation of this insightful thought in a classroom setting.

Apart from intelligibility and compressibility framework, in Chapter 23, Graeme Couper reviewed major theories on language learning in the fields of applied linguistics (SLA theories), educational psychology (social theories of learning), phonology and L2 speech research and

cognitive linguistics and phonology (a pronunciation learning and teaching framework). In addition to the review, Couper provides five practical teaching tips derived from the theories for English pronunciation in class. Readers who are interested in obtaining knowledge on language learning theories across the four different areas of studies would benefit from perusing the first part of the chapter and language teachers who strive for best practice of teaching pronunciation would enjoy reading the second part.

In Chapter 24, Robin Walker and Wafa Zoghbor address teaching English pronunciation from a perspective of an English as a Lingua Franca (ELF); Lingua Franca Core (LFC) (Jenkins, 2000), significant segmental and suprasegmental pronunciation features that needed for intelligible oral communication between NNS English speakers is discussed as well. After delineating the concept of ELF and LFC, the authors devote much room to consider practical aspects (e.g., model selection, techniques) of practicing ELF approach in classroom. This chapter would provide a stunning opportunity for EFL teachers and leaners who have been skeptical about foreign accent in English.

In Chapter 25, Marnie Reed and Christina Michaud discuss why teaching intonation is pivotal in an aspect of pragmatic function and they examined intonation learning activities in currently used teaching materials and textbooks. In addition, based on research findings, they argue that production-oriented intonation teaching, which is a current approach to intonation teaching, is not only insufficient but also undesirable; to be more successful in teaching intonation, they suggest teachers encourage their students to engage in metacognition while conducting production-oriented activities. This chapter ends with five recommendations for teaching intonation through metacognition.

In Chapter 26, in accordance with authors of the preceding chapters, Laura Sicola and Isabelle Darcy emphasize the importance of pronunciation teaching in L2 education; however, their claim centers on the integration of pronunciation into other areas of language teaching such as grammar and even other subject-matter courses under the communicative approach -- the effects of pronunciation teaching would be maximized when both form and meaning are addressed at the same time. Despite the fact that their argument sounds quite ideal from the perspective of form-focused communicative language teaching, recommendations that they provide regarding how to incorporate the pronunciation into other courses do not seem either innovative in terms of other areas of language teaching or concrete enough with regards to content courses for language teachers to endorse and practice them in class.

In Chapter 27, Wayne Dickerson discusses the usefulness of English orthography as an assisting tool to promote language learners' speaking skills from a segmental to rhythmic level; even though a single sound does not correspond to a single letter, he argues that there are rules or regularities in spelling that govern pronunciation system in English. Thus, if learners learn these rules, they would make sound predictions about pronunciation of both segmental and suprasegmental features. Dickerson, on the one hand, shows how English pronunciation can be accurately predicted based on orthography with various examples, which would look revealing to

those who persistently believe that it is impossible to make a one-to-one connection between sounds and spelling in English. On the other hand, the rules do not deal with all possible instances of pronunciation nor empirical evidence is provided to show the effectiveness of this approach in pronunciation education.

Quite differently from the other chapters, in the last chapter, Rebecca Hincks overviews the development and use of technology in the field of pronunciation teaching and learning for various purposes: speech training, assessing, and conversation practices (e.g., computer-assisted pronunciation training). In the overview, she also addresses limitations of using contemporary technologies for specific purposes in the field. This chapter is highly recommendable for language teachers, who aim to incorporate current technologies to pronunciation teaching in their instruction, to read so that they can maximally utilize the advantages that technology provides, acknowledging the limitations.

The section of the handbook has both weaknesses and strengths. Most of all, covering all three aspects of pronunciation teaching from theories, and research to classroom practices might be a double-edged sword. It meets, on the one hand, a wide range of readers' -- theory-, research- and teaching-oriented audience -- needs whose aim is to gain knowledge on pronunciation teaching with a particular orientation. On the other hand, the depth of discussion seems neither broad nor deep enough in all three directions, especially practical aspects of pronunciation teaching in a classroom setting. It seems that all the chapters gear toward theoretical and/or research aspect. By extension, most of the chapters discuss the topic, pronunciation teaching with many discipline specific terminology (for instance, especially, Chapter 23); thus, it would be expected to be challenging for the readers who do not have professional knowledge on the field, to some extent, to thoroughly comprehend the chapters and to in turn benefit from the reading. This is clearly opposed to what is stated in the introduction, "The Handbook is intended for applied linguistics and for teachers, for who are experts and for those who are not" (p.xii). Last but not least, it is surprising that no single chapter in this section is devoted to thoroughly discussing a core aspect of pronunciation teaching, assessment of pronunciation -- how to assess pronunciation through tasks in classrooms (e.g., see Part 6 in Brown, 1998; Chapter 8 Celce-Murcia, Brinton, Goodwin & Griner, 2010 for references) and what issues should be considered (e.g., Levis, 2006) -- despite the fact that assessment is not separable to language learning and teaching. However, these shortcomings should not discourage prospective readers from reading the section and the book itself in that they provide insightful and comprehensive discussion of experts on essential aspects of pronunciation teaching in English. This strength definitely overrides the weaknesses.

REFERENCES

Brown, J. D. (1998). *New ways of classroom assessment. New ways in TESOL series II. innovative classroom techniques*. Bloomington, IL: Pantagraph Printing.

264

Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (2010). *Teaching pronunciation hardback with audio CDs (2): A course book and reference guide*. New York, NY: Cambridge University Press.

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

Levis, J. M. (2006). Pronunciation and the assessment of spoken language. In R. Hughes (Ed.), *Spoken English, TESOL and applied linguistics: challenges for theory and practice* (pp.245-270). London, UK: Palgrave macmillan.

MOBILE APP REVIEW

English File Pronunciation App
Joseph Geluso, Iowa State University

INTRODUCTION

English File Pronunciation (EFP) is a software application (app) (2012) published by Oxford University Press. The EFP app is intended for use on mobile devices such as smartphones and tablets via Apple's iOS or Google's Android operating systems. The app will cost consumers USD $5.99 on both Apple's app store and Google Play for the Android version. For this review, the iPad version of the EFP app was used. The EFP app is marketed as supplementary content to Oxford's *English File Student's Book 3rd edition* (Oxemdem & Latham-Koenig, 2015), and content from the app should be familiar to those who use the textbooks. The app is designed to give users a method to "Practise sounds, words and sentences" as the app's tagline declares: "The app that gets you talking". Oxford attempts to entice consumers by claiming that users will "see your communication and pronunciation improve with the unique English File Sounds Chart."

OVERVIEW OF APP

When you first open the app you are prompted to select "American English" or "British English" as shown below in Figure 1. This setting can be changed at any time from the home screen by clicking the gear in the upper right-hand corner.

Figure 1. Initial screen of EFP app offering choice of American or British English.

The main component of the app is the "English File Sound Chart". Your first encounter with the chart will feature instructions that explain the interface and how to use it. The interface itself is quite intuitive and frankly could be figured out with a few minutes of experimentation by anyone familiar with the ubiquitous touchscreen interfaces of most modern smartphones. The sound chart that makes up the backbone of the EFP app is depicted below in Figure 2.
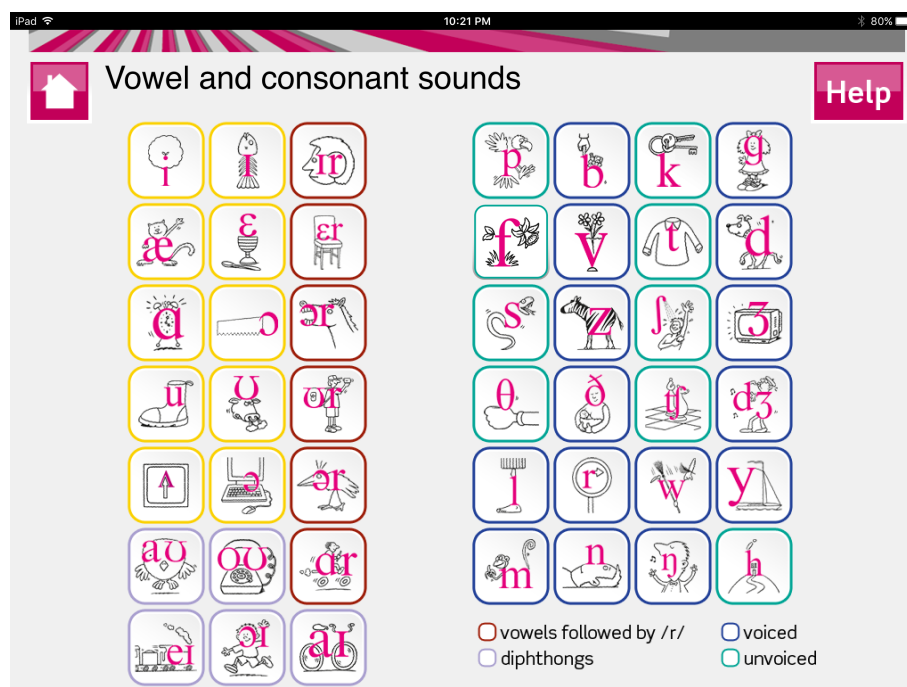
Figure 2. Sound chart

As can be seen in Figure 2, the chart uses characters from the International Phonetic Alphabet (IPA) to label the individual tiles that represent phonemes. Each tile for the various phonemes is surrounded by a red, light purple, blue, or aqua-colored border. The colors represent "vowels followed by /r/", "dipthongs", "voiced", and "unvoiced", respectively. Tapping a tile foregrounds it and allows the user to hear the phoneme pronounced in isolation, and is followed by an example word featuring the target phoneme. So, for example, after pressing the /s/, the tile will enlarge and the user will hear "S. Snake." (see Figure 3).
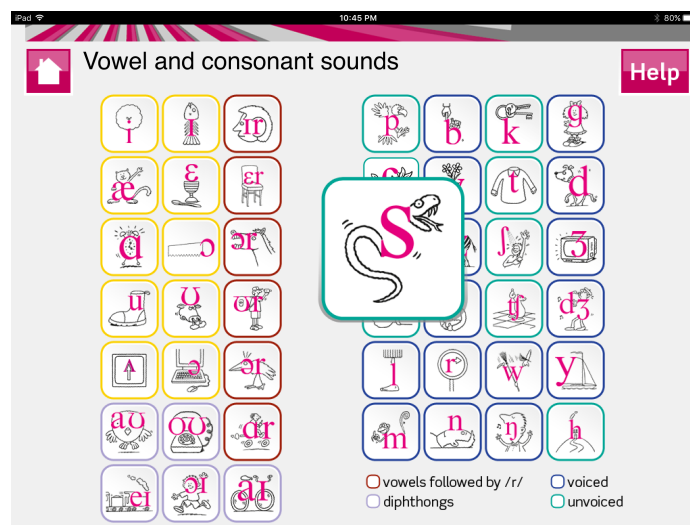
Figure 3. Tapping a tile enlarges it and plays a sample recording

Tapping on a tile a second time takes the user to a new screen that features more sample words that contain the phoneme. Users can play a recording to hear the phoneme in the context of each of the words. The user can also choose to view a set of three sentences containing the target phoneme. In addition to being able to listen to the words or sentences, users can record their own voice and play it back to compare with the example pronunciation provided by the app (see Figures 4 and 5).
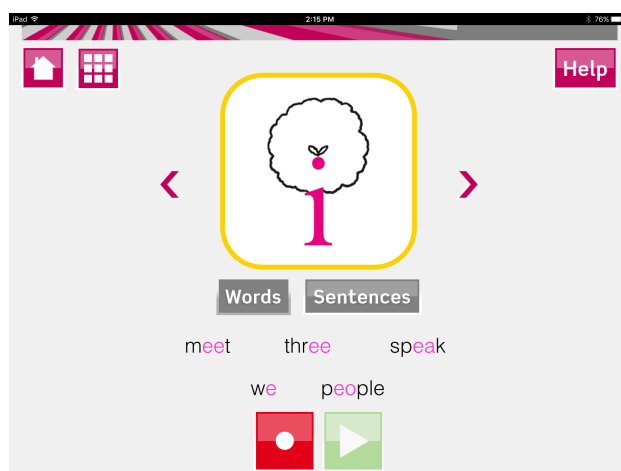

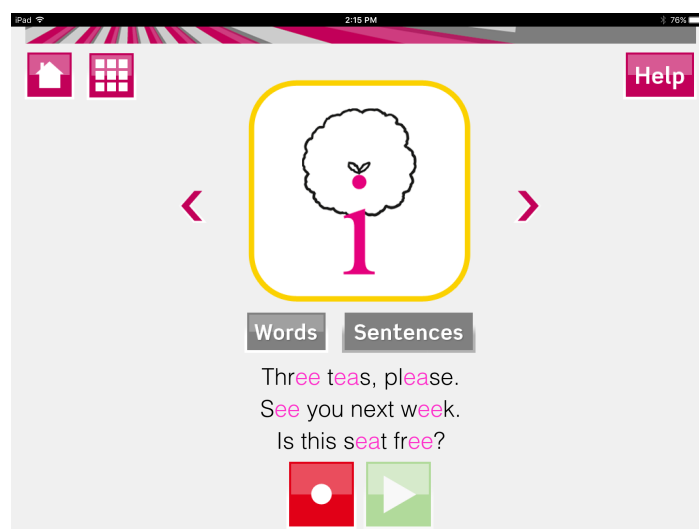*Figure 4.* Words for /i/ and record and playback buttons

Figure 5. Sentences for /i/ and record and playback buttons

The second major component of the app is a matching game that can be played with or without audio. Without audio, the user is instructed to match one of two phoneme options to a given word (see Figure 6). Users can also play the game with audio, in which case a word containing the target phoneme is uttered and the user has to choose the tile that contains the matching phoneme (see Figure 7). Users can listen to the audio as many times as they like.
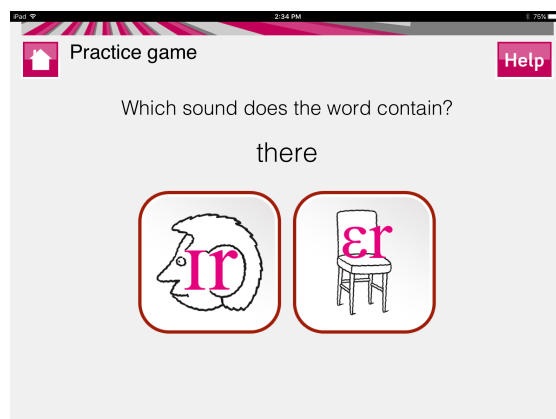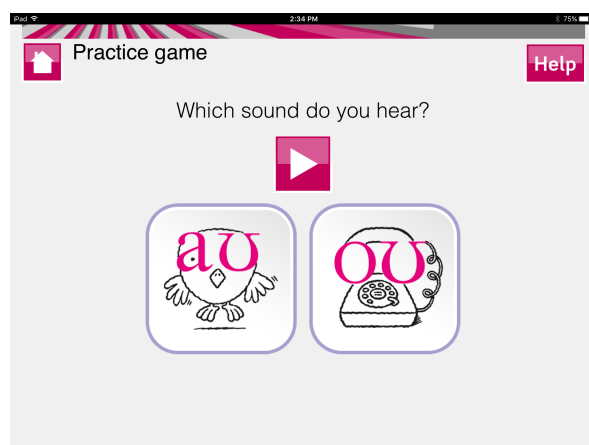

Figure 6. Matching game without audio

Figure 7. Matching game with audio

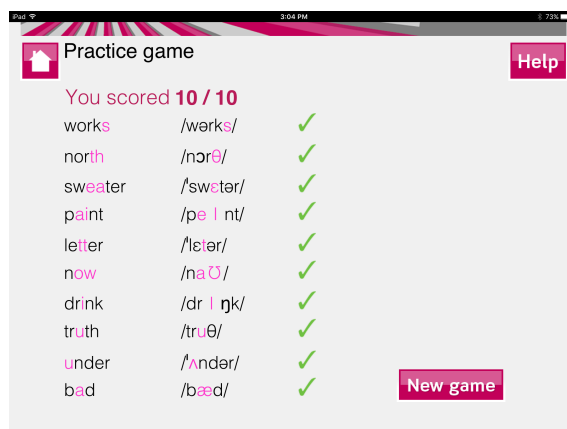When completed, the app presents the user with his or her final score (see Figure 8).

Figure 8. Feedback from matching game

EVALUATION

As is seen in the various screenshots throughout this review, the EFP app uses a variety of colors and pictures to make its interface visually stimulating and draw the user's attention to where target phonemes lie. This visually enhanced input is in line with Chapelle's (2003) suggestions to promote second language learning in CALL environments (p. 38). Beyond visual enhancement, the app also utilizes aural enhancement to highlight target phonemes. However, as Chapelle notes, the principle of enhancing input does not work as neatly with aural content as it does with written content (p. 41). Indeed, the in-app audio example words and sentences stress the target phonemes so much that it sometimes results in an overemphasis and unnatural distribution of stress in the word. This is especially true in the case of exercises targeting vowels.

Like Chapelle's (2003) call for grounding CALL in SLA theory, Levis (2007) argues for the same with respect to computer assisted pronunciation teaching (CAPT). This app does not always demonstrate a firm grounding in theory, and at times falls victim to the same trap that ensnares so many CALL and CAPT applications in that it "look[s] suspiciously like traditional, drill-oriented pedagogy in new clothing" (Levis, 2007, p. 185). To escape situations like this, Levis points to Pennington's (1999) CAPT design principles. Pennington maintains that CAPT should start from a theoretical position, and offers 10 suggestions for improving CAPT pedagogy. Among these are: *set an overall goal performance; build specific targets for performance; link pronunciation to other learning and communicative goals; design on a principled curriculum;* and *provide for exploration of database.* While the EFP app arguably meets the last two suggestions, it does nothing with respect to the first three. Additionally, as Levis points out, feedback is an area where CAPT systems routinely fall short as they are unable to perform an automatic and accurate diagnosis of pronunciation. In fact, for the purposes of this review I invited a number of users to play with the app for a short period of time and a weakness many of them pointed out was lack of feedback.

With respect to the matching game, while fun for a while, it fails to deliver elements associated with successful digital games-based learning. Elements such as competition with others and rewards for achievement (Intratat, 2011; Prensky, 2003) could be introduced by users who take it

upon themselves to enrich the experience, but they are not incorporated into the build of the app. In fact, Oxford (2015) has games such as *Stress Monsters* on its companion website to the English File textbooks that are more engaging and better adhere to Prensky and Intratat's elements of successful digital games for learning than the app in question does.

All said, the app is successful on a number of fronts. For example, the ability of the users to be able to record their own voices to compare to the sample pronunciation provided in the app is a nice feature. Casual users will likely be entertained by this feature and also by the short matching games in the app. As Levis (2007) notes, computer assisted pronunciation teaching (CAPT) seems to work well with children and this app would likely be enjoyed by a younger audience with its bright visuals and interactive features. Ultimately, however, I do not believe that the content justifies the $5.99 price tag. The app in its current state would be better packaged as free downloadable content for those who bought the English File textbooks.

REFERENCES

Chapelle, C. (2003). *English language learning and technology. Lectures on applied linguistics in the age of information and communication technology.* Amsterdam: John Benjamins.

Intratat, C. (2011). Alternatives for making language learning games more appealing for self-access learning. *Studies in Self-Access Learning Journal: Special issue on CALL, E-Learning, and M-Learning, 2*(3), 136-152.

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

Pennington, M. (1999). Computer-aided pronunciation pedagogy: Promise, limitations, directions. *Computer Assisted Language Learning, 12*(5), 427-440.

Prensky, M. (2003). Digital game-based learning. *Computers in Entertainment, 1*(1), 21.

Oxemdem, C. & Latham-Koenig, C. (2015). *New English File* (3rd edition). New York: Oxford University Press.

Oxford. (2012). English File Pronunciation App. Retrieved November 15, 2015 from https://itunes.apple.com/us/app/english-file-pronunciation/id520767531?mt=8

Oxford. (2015). Retrieved December 4, 2015, from https://elt.oup.com/student/englishfile/elementary3/c_pronunciation/ef_stressgame?cc=us&selLanguage=en

SOFTWARE REVIEW

ToPhonetics
Rosalie Hirch, Iowa State University

INTRODUCTION

ToPhonetics is a program that takes any English script and automatically converts it into another script such as IPA. The program began as a free-access website available through lingorado.com, a Russian website dedicated to learning English, particularly pronunciation. A mobile application (app) based on the web program was developed in 2013, and is currently available on both iPhone ($3.99) and Android ($2.20). The iPhone app is reviewed in this article, with reference to the original website.

ToPhonetics is intended to perform as computer assisted pronunciation training (CAPT), to help English learners improve their pronunciation. There has been some suggestion that mobile CAPT apps may be particularly useful for pronunciation training, though little research exists in this area (Thomson, 2011). CAPT software can be useful for students in an EFL context, where learners may have minimal access to native speaker input (Eskenazi, 1999). Other potential benefits of CAPT for ESL learners are assisting with spelling and pronunciation (Finnegan, 2004) and improving listening comprehension (Thomson, 2011). One final helpful element is the use of IPA as a visual aide, as visualizations have been demonstrated to help students with pronunciation (Godwin-Jones, 2009; Thomson, 2011). The following review briefly introduces the *ToPhonetics* website (Jans, 2013) to show the ESL focus, then goes into greater detail about the app before giving an evaluation.

PRODUCT DESCRIPTION

The program on the website is a simplified version of the app in that it has far fewer features, but demonstrates the intended use for ESL students. Figure 1 shows options at the top for various languages, but contrary to what users may expect, these simply give the instructions in other languages; the program will not convert foreign words. The IPA text that is produced is the same as that produced on the app, and is comparable to what other online transcription programs produce. In this sense, the website produces nothing original. The unique elements are in the app.
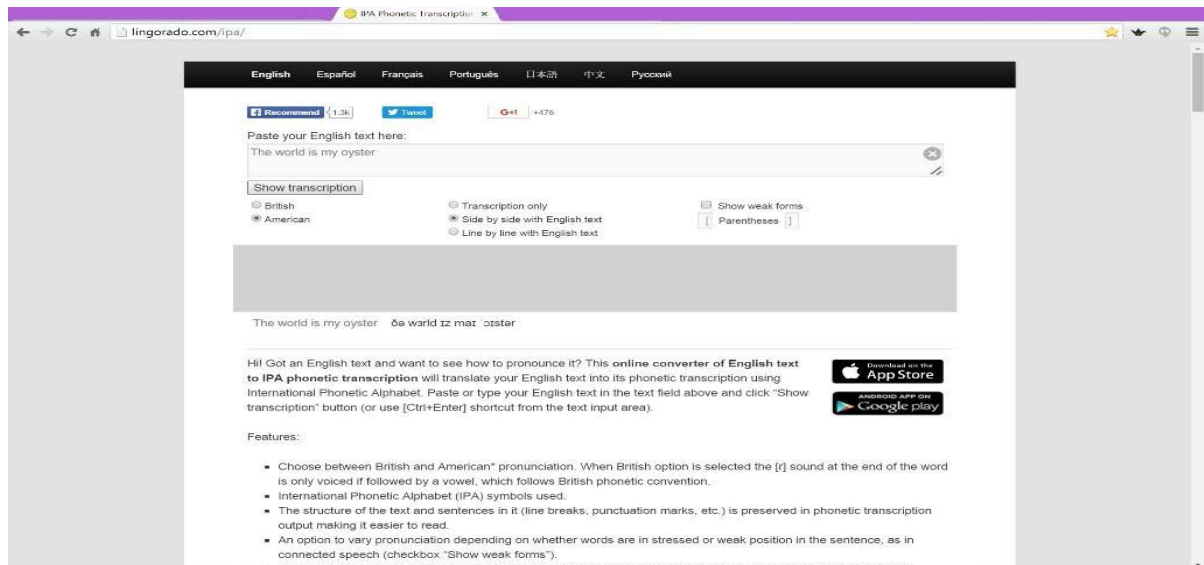
Figure 1. ToPhonetics website

MOBILE APPLICATION

As soon as users open the app, they see a space for text entry, as shown in Figure 2. The interface is straightforward and intuitive, since the keyboard is identical to that of other apps incorporating text. *ToPhonetics* takes advantage of autocorrect and accepts dictation by tapping the microphone. It is important to note that using dictation does not represent the speaker's phonetic production, but "normalizes" the text with standard orthography and translates that text. It is also unclear how this function would work with foreign accents.

Figure 2. ToPhonetics app homepage.

The "X" in the upper right corner clears the text that has been entered, while the gear symbol in the upper left corner opens the settings. The options for settings are shown in Figure 3; explaining these shows the app's features.

The first option is "English dialect", which gives a choice between American or British English, with British English being the default (the app was originally designed for a European audience). The second option, "Show weak forms", is intended to represent some suprasegmental aspects, specifically the difference between "normal" and contextual pronunciation. So, for example, /to get/ is [tu gɛt] in regular mode, but becomes [tə gɛt] when weak form is selected.

Figure 3. ToPhonetics app options.

The next option is "Transcription character set", the choices for which are IPA standard (the default setting), IPA basic, Russian, and Katakana (Japanese). Although the app offers other language options, the developers explicitly recommend against using these other languages, which fits research into problems with using first language orthography (Godwin-Jones, 2009). "Include original text" refers to the layout of the output. Original text can be presented above the IPA script, which can be effective for learners to improve pronunciation (Eskenazi, 1999). The "Parentheses" option gives users the ability to separate each word for easier reading.

The "Colours" option is also intended to make interpretation easier by color-coding the words in the IPA transcription. Figure 4 shows the different presentation options available. "Multiple pronunciations" means that words in blue have several options for pronunciation, while "Not found" refers to words that are not in the library. The others are fairly intuitive, except for "Cursor", which is used for the speech option (explained below). These visualizations can be helpful for learners as they can draw a learner's attention to aspects of pronunciation; such visual cues have been demonstrated to improve pronunciation (Thomson, 2011).

"Speech rate" refers to an option on the app to listen to the text, described below; it is enough to say now that the speech rate for that option can be adjusted here. The final section of the options is "Feedback", which is how users can give feedback to the developers through several platforms. Feedback on learners' pronunciation is not offered on this app.

Figure 4. ToPhonetics app colors.

Once options are set, the user can enter text on the home page. An example of output is shown in Figure 5. In addition to the initial output presented on the screen, the user can double click any word to get the dictionary definition from the Oxford American Dictionary; the definition page also has an option to do an automatic Google search. In the case of words that are in blue, such as "bought" in Figure 5 above, double clicking will give both pronunciations; for /bought/, these are [bɑt] and [bɔt]. The file symbol at the top right corner gives various standard options for sharing and saving the output.

Figure 5. ToPhonetics app output: minimal pairs.

Tapping the arrow in the bottom left corner plays the text as read by a computer voice recording. The green cursor mentioned above highlights the text as it is being read, as shown in Figure 6. *ToPhonetics* has made some attempts to add suprasegmental elements to the recordings, specifically intonation. For the texts in Figure 6, for example, the first question is read with a downward final inflection, while the second question is read with an upward inflection, appropriate for those questions. Though the readings definitely sound like a computer and therefore not very authentic, they do try to give the impression of appropriate intonation.

Figure 6. ToPhonetics app speech function.

EVALUATION

This app focuses almost entirely on segmental aspects of pronunciation, which can be effective for beginners who have little experience with differences in pronunciation between their L1 and English (Eskenazi, 1999). Interestingly, comments left by users on the website seem to indicate that advanced learners also frequently make use of the app (Jans, 2013). This suggests that *ToPhonetics* may be effective as a resource for English language learners generally. It could also encourage autonomous learning (Lu, 2010; Thomson, 2011). In particular, students have the ability to move at their own pace and focus on aspects of pronunciation that they find most difficult. The app also allows students to specialize materials to their specific needs such as discipline-specific vocabulary (Godwin-Jones, 2009). All of these are effective for drawing attention to pronunciation at the segmental level (Thomson, 2011; Munro, 2005).

However, there are some limitations, particularly at the level of prosody. Though research indicates that beginners gain knowledge from segmental focus, there is also evidence that beginners can benefit from focus on prosodic elements early on (Eskenazi, 1999; Godwin-Jones, 2009; Munro, 2005). Unfortunately, while the app does make attempts to represent prosodic elements such as stress and intonation, these elements are difficult to interpret in the app and are not represented visually; learners' attention is therefore not drawn to suprasegmental elements, making those aspects less effective.

Another element important to improving pronunciation is variation; English language learners exposed to speech from a variety of native speakers and dialects usually improve

more quickly than those with input from only one speaker or dialect. Access to multiple voices should be a benefit of CAPT (Thomson, 2011). An app like *ToPhonetics* does offer one voice and could be helpful in remote places where learners have few opportunities to interact with native English speakers. Unfortunately, the computer voice is not very natural, so as input it is less than ideal. One final element that would be helpful is feedback for students. This could come in many forms such as recording students' voices or providing visualizations of their speech for comparison purposes (Thomson, 2011; Lu, 2010; Godwin-Jones, 2009). Without such feedback, it is difficult for students to objectively identify their own pronunciation problems, perhaps rendering the app a better resource material than training tool.

Overall, one of the fundamental issues for this app, as with almost any CALL technology, is that training is key (Chapelle, 2003; Eskenazi, 1999). Incorporating an app like this into a course or self-study requires careful research and consideration into how it can best be utilized. It should also be emphasized that the pronunciation information offered in the app is prescriptive, not descriptive, an aspect that needs to be explained to users. For this reason, the app may also prove useful for linguistics students learning about prescriptive versus descriptive pronunciation.

SUMMARY

ToPhonetics is a useful tool and easy-to-use resource for helping students with segmental aspects of pronunciation. In particular, it can give basic information on pronunciation that will be useful for beginning learners, and could be a reliable resource for more advanced learners as well as linguistics students. However, the app should be used carefully; students in particular need to understand the limitations of the platform and should be sufficiently trained before using it.

REFERENCES

Chapelle, Carol A.(Author). English Language Learning and Technology. Lectures on applied linguistics in the age of information and communication technology. Philadelphia, PA, USA: John Benjamins Publishing Company, 2003. http://site.ebrary.com/lib/modarr. (2003), (March).

Eskenazi, M. (1999). Using a Computer in Foreign Language Pronunciation Training : What Advantages ? *CALICO Journal*, *16*(3), 447–470.

Finegan, E. (2015). *Language: Its Structure and Use. 7th Edition*. Cengage Learning: Stamford CT.

Godwin-Jones, R. (2009). Emerging technologies: Speech tools and technologies. *Language Learning & Technology*, *13*(3), 4–11. Retrieved from http://llt.msu.edu/vol13num3/emerging.pdf

Jans, D. (2013). Lingorado. Retrieved December 14, 2015, from http://www.lingorado.com

Lu, D. (2010). A salutary lesson from a computer-based self-access language learning

project. *Computer Assisted Language Learning*, *23*(July 2014), 343–359. http://doi.org/10.1080/09588221.2010.511588

Munro, J. (2005). and Pronunciation Teaching : A Research-Based Approach, *39*(3), 379–397.

Thomson, R. (2011). Computer Assisted Pronunciation Training : Target- ing Second Language Vowel Perception Improves Pronunciation. *CALICO Journal*, *28*(3), 744–765.

Lee, E. (2016). Pronunciation in EFL instruction: A research-based approach [Review]. 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. 281-284). Ames, IA: Iowa State University.

BOOK REVIEW

Szpyra, J. (2015). *Pronunciation in EFL instruction: A research-based approach*. Multilingual Matters.

Elizabeth Lee, Iowa State University

This book addresses the importance of the pronunciation in EFL education and begins with a discussion on the the theory of EFL pronunciation instruction before moving on to instructional techniques that pronunciation instructors could use to teach English pronunciation in an EFL context. Each chapter is divided into two parts A and B: part A gives an overall theoretical picture of topics related to pronunciation teaching, and part B summarizes empirical studies that would support some of the arguments made in part A. As the author mentions in her preface, the book is for current and prospective EFL and teacher trainers interested in improving their teaching skills and pronunciation instruction, pronunciation specialists and students of applied linguistics (Szpyra, 2015, pp. viii-ix). While the author attempts to reduce the amount of technical words, readers who have very little knowledge in pronunciation, phonetics and phonology, would find some of the terminologies difficult to comprehend without referring to outside sources. Nevertheless, I think this is a good introduction to any person who is interested in either teaching or conducting research in pronunciation teaching. This book review would only focus on the contents of the first two chapters of the book.

The author begins with the importance of teaching pronunciation: she argues that pronunciation promotes listener's positive perception of the speaker, good oral communication, and increases speaker's confidence to speak good English (pp. 2-4). However, the author notes that teaching pronunciation has often been neglected because many teachers find pronunciation teaching difficult and least useful; and there are not many teaching materials that support language teachers' ability to teach pronunciation (pp. 4-6). Szpyra thinks teaching pronunciation is important and that rather than emphasizing native-like pronunciation, it is more important to promote "comfortable intelligibility" (7), where the speaker would have to speak intelligibly enough for the listener to understand in a communicative setting.

While various approaches to teaching pronunciation have been advocated, the author believes that the best model for teaching pronunciation in EFL context is one that combines both pronunciation models of English as a Lingua Franca (ELF) and English as a Foreign Language (EFL): the New English as a Lingua Franca (NELF) (p. 25). This approach would allow language learners to receive both native and non-native models of English pronunciation, and recognize the importance of being exposed to various types of pronunciation, and practice the kind of pronunciation that would appear most intelligible to native and non-native speakers of English.

281

The author then moves on to discuss the difference between EFL and ESL, arguing that ESL learners are more likely than EFL learners to be exposed to English outside of the classroom setting, and therefore have more hours and opportunities to practice the targeted L2 pronunciation. EFL learners, on the other hand, are more likely to have limited English exposure, within a classroom setting, and they may or may not have access to communicating with native speakers of English. As such, phonemic exercises that may be suitable in an ESL context may not transfer easily in an EFL context without considering the different teaching and learning settings.

The author recommends that EFL teachers should self-examine their pedagogical practices and evaluate whether their phonodidactic approach is serving the needs of his or her students (44). For example, a teacher would ask herself or himself, what major pronunciation problems are experienced by EFL learners, and therefore which pronunciation techniques and learning tasks she or he should provide to address learners' pronunciation errors. Importantly, the teacher should consider the learning context and what consequences may follow for the learner who is speaking in L2.

Following her theoretical rationale for implementing a NELF approach to teaching English pronunciation in an EFL context, she provides a few studies she had done that would support her rationale for teaching pronunciation. In the author's own research, she found that listeners' attitudes toward speakers whose speech were heavily accented and less comprehensible, were less favorably evaluated than speakers whose speech were less accented and more comprehensible (48). Her study confirms previous studies that suggest that pronunciation and level of accentedness influence listeners' perceptions of speakers' personality and competency, and so pronunciation would be an important feature to be taught in EFL.

Based on a survey she had conducted on EFL students and teachers in Poland, it appears that pronunciation is not as emphasized as heavily as other areas of language learning. Phonetic training is usually practiced with beginning learners of English but less so with intermediate or advanced learners of English. In classroom settings, a large proportion of participants indicated that they learn pronunciation mostly through teacher correction or by listening to audio recordings. While it would not be fair to assume that every teacher does not prioritize pronunciation when teaching English, the results from the questionnaire seemed to suggest that pronunciation and other phonetic and phonemic trainings are not at the forefront of EFL education. The author may have included these studies in the first chapter to point to the lack of research and support for good pronunciation teaching.

In the second chapter, Szpyra considers which phonetic features should be focused in order to help learners produce comfortable intelligibility. She lists teaching high functional load (those where you would see a lot of minimal pair contrasts); words that would have "high return" (i.e. speakers giving lower effort in order to produce higher result); and words that would match the learners' end purpose for studying English (pp. 68-71).

282

Again, the author emphasizes the importance of intelligbility in oral communication and considers the different proposal for English pronunciation priorities: Jenkins' Lingua Franca Core, Amalgam of English and International English, and Collins and Mees' Error Ranking (88). While proposals vary in terms of which pronunciation to prioritize, focus on consonants, vowels, and word stress patterns appeared to be common priorities found in all four models. A further assumption that is made in all four models is that pronunciation errors are universal. The author notes that none of the proposals are research-based, and so what these proposals consider pronunciation priorities may not be as reliable as one would deem. In other words, without strong empirical research to back up their arguments, it is difficult to validate which pronunciation priorities are more important than others.

Instead of listing which consonants, vowels, and prosody to focus on, the author of this book suggests that teachers consider what common L2 pronunciation errors their learners make. The author recognizes that students' L1 can influence how they acquire L2 pronunciation, and therefore teachers may need to prioritize different pronunciation features depending on learners' L1. Moreover, the author thinks rather than focusing solely on phonetics, more emphasis should be given to pronunciation of phonetically difficult words (90).

In addition to helping learners pronounce phonetically difficult words, the author also suggests teachers addressing local errors, which are perceived to be more severe than global errors. According to the author, global errors have to do speaking English with a foreign accent due to L1 phonological and phonetic transfer; and local errors are mispronunciations of individual words that are considered more severe and more difficult to fix than global errors (93). The author reasons that local errors are severe because there are interference from L1 and L2 that may hinder learners' ability to pronounce L2 words properly. The author is likely to promote L2 pronunciation teachers who prioritize common local errors found in L2 learners' speech in order to correct their mispronunciations, rather than those relying on a model that focuses on different consonantal, vocalic, or prosodic features without considering the needs of their learners. Although this is not to say that the author opposes phonetic training, the author seems to suggest that current pronunciation teaching favors phonetic training over other pronunciation activities; and in this chapter, the author clearly points to the importance of attending to teaching phonetically difficult words as well as addressing learners' local errors.

The author then argues that the debate of whether to prioritize segmental vs suprasegmental issues, in order to improve learners' intelligibility, is not as important as figuring out to what extent segmental and suprasegmental features are similar and different between students' L1 and L2. If L1 and L2 were found to have similar segmental features but differing suprasegmental features, then the teacher should focus more on suprasegmentals; and if L1 and L2 were found to have similar suprasegmental features but differing segmental features, then the teacher should focus more on segmentals.

In other words, learners' pronunciation needs may vary depending on their L1 backgrounds, and so, assuming that one feature is more important than another is suggesting that L1s have similar

283

pronunciation features and therefore teachers simply need to focus on either segmental or suprasegmental features. However the author would think this is not true as certain languages share more segmental features with English, and other languages share more suprasegmental features. Still, for beginning learners of English, it would benefit to focus more on segmental issues, and for more advanced learners of English, it would be useful to concentrate on suprasegmental issues.

In part B, the author conducts research about intelligibility and pronunciation priorities of Polish learners of English. When Polish learners of English were asked to evaluate the intelligibility of two different non-native speakers of English, her study found that intelligibility is correlated with annoyance (119). That is, a speaker who produced global errors (i.e., prosodic issues) was perceived to be less annoying than a speaker who produced local errors (i.e., idiosyncratically deviant words). The author concludes that pronunciation instruction should focus on not only segmental and suprasegmentals features but also whole words that are commonly mispronounced by EFL learners.

As the author notes in part A of chapter 2, pronunciation priorities largely depend on learners' L1. In the case of Polish learners of English, she found that learners frequently struggle with pronouncing English words that have a different spelling to pronunciation rule from Polish. In addition, cognates and words with difficult stress patterns were also found to be problematic for these learners. If one were to teach Polish learners of English, pronunciation teachers would have to address such issues in order to help learners improve their English pronunciation.

As a person with limited knowledge of pronunciation teaching, I've gained a lot of insight about the current issues that are going on with pronunciation teaching in EFL context (especially in Poland). The first chapter discusses how pronunciation has been taught in various English language learning contexts. In the second chapter, the author talks about pronunciation priorities. While the author did point out the limitations of her study being limited to a Polish EFL context, I think this book would attract a wider audience if the author had done meta-analyses of current pronunciation teaching studies in other EFL contexts. Despite the lack of generalizability of her studies, I appreciate that the author made great efforts to cover a host of pronunciation-related issues in concrete detail, while still maintaining her main argument, which is, helping learners produce intelligible speech.

284

SOFTWARE REVIEW

Virtual Talking Head
Ivana Lucic, Iowa State University

INTRODUCTION

Animation could play a potential role in improving the human learning process, especially in promoting deep understanding of the subject matter (Ahmah Zamzuri, 2013). Using more than one modality in learning enables the creation of referential connections, which facilitates learning (Zhu, Fung & Wang, 2012). The speech mapping concept has become more influential in the ways that visual modeling is done since the early 1990s. It is viewed as a useful framework for pronunciation training that provides a visual display of a spectrum of sounds produced by a particular speaker. It enables better understanding of the link between speech production and speech perception (Abry & Badin, 1996).

The three-dimensional virtual talking head was developed as a virtual anthropomorphic robot based on physical modelling of the articulatory, aerodynamic and acoustic phenomena involved in the audio-visual production of speech (Badin, Bailly & Boe, 1998). This specific model was developed in Grenoble, France at a CNRS (Centre National de la Recherche Scientifique) research unit. Even though this model is not publicly available, concepts and frameworks provided by these researchers could be used for developing a similar model to aid second language speakers in their learning of pronunciation. CNRS created the Virtual Talking Head to manipulate audio-visual speech stimuli in order to fulfill two main tasks:

(1) Evaluating and improving the learner's perception of the target language sounds,

(2) Helping the learner produce the corresponding articulations by acquiring the internalization of the relations between articulatory gestures and resulting sounds (Badin et al., 1998).

An L2 learner can be considered phonologically deaf in regard to particular sound categories, which means they are not able to distinguish speech sounds that do not belong to the phonological inventory of their L1, or they are not similar enough to the existing sound map. If there is a perceptive issue with sound recognition, production as well will most likely be problematic (Badin et al., 1998). In addition, existing research related to psychology and neuroscience shows that speech production and speech perception occur in separate paths in human brains (Skoyles, 2010). This means that examining and analyzing both speech production and speech perception as complementary skills could lead to improvement of the process of pronunciation teaching in language classrooms (Badin et al., 1998). In order for the learner to grasp proper production of perceptively

acquired sounds "they must shape their vocal tract and dynamically coordinate articulators to produce these specific acoustic targets by means of maneuvers that may be new to him/her" (Badin et al., 1998, p.1). This is similar to any type of muscle exercising: the vocal tract consists of many muscles that need to learn how to move in a different manner, and the more practice the learner provides for it, the better the acquisition and production of sounds.

HOW IT WORKS

To further explain the importance of both production and perception form pronunciation improvements, Badin et al. (1998) quotes LeBel who said that three of the "grands moyens [big means]" in the domain of phonetic correction are directly related to perception and production:

> (1) Auditory discrimination (one can pronounce well only what one can perceive well),

> (2) Articulatory and acoustic composition (the learning process will be more efficient if the learner knows which articulator he/she should pay attention to in order to correct a specific problem),

> (3) Combinatory phonetics (various coarticulation effects can be used to induce the right articulatory gestures for a given phoneme).

The space of articulator's positions, the geometric and the acoustic/auditory space, and the relations between them are implemented in a virtual talking head, which is an anthropomorphic model of speech production.

In order for the virtual talking head concept to function accurately, it was necessary to obtain complementary data from various experimental setups for various reference subjects. The subjects involved had to produce the same speech material in the same (controlled) conditions. In this way, the framework provided complete and accurate representations of the different mechanisms involved at different levels in the speech production chain and at constructing a comprehensive model (Badin et al., 1998). Some of the most important methods used to reach a valuable level of accuracy are:

> "(1) Cineradiography that produces limited but extremely valuable sets of midsagittal vocal tract contours,

> (2) Pneumotachometry that provides air flow at the lips and intraoral pressure, video labiometry that furnishes a geometric description of lips from front and profile views,

> (3) Electromagnetic articulometry that delivers the X/Y coordinates in the midsagittal plane of a few points attached the tongue or to the jaw,

> (4) Magnetic Resonance Imaging that results in full 3D geometric descriptions of sustained articulations" (Engwall, 2003).

The experimental setups consisted of multiple tests that needed to be performed on participating subjects, including MRI (magnetic resonance imaging), EPG (electropalatography) and EMA (electromagnetic articulography). "The shape and parameters are determined through statistical analysis of static MRI data, the parameter activation is based on the combination of MRI and EPG, and the timing of the movements is determined from EMA data" (Engwall, 2003, p. 312). Even though such creation of the model ensures reliability in recreation of virtual model sounds, the conclusion Engwall (2003) came to is that the static MRI data needed to be complemented with real-time data, in order to generate a model that is fully representative of running speech. This makes the model somewhat inconvenient. Nevertheless, the three-dimensionality of the framework provides learners with a more accurate representation of the inner processes, as it enables them to visualize how the vocal tract works when producing specific sounds.

USEFULNESS IN TEACHING

The main tasks of a teacher who uses the virtual talking head framework is to both evaluate and improve the learner's ability to perceive the vowels and consonants of the target language. Elaborating teaching strategies is another way the virtual talking head could positively influence the learner – teachers have a chance to help them find and understand the right articulatory gestures to produce what they learned to perceive (Badin et al.,1998). The virtual talking head framework can be used for generating of appropriate stimuli in order to evaluate the learner's ability to discriminate sounds in the target language, and to progressively improve the relationship between their productive and perceptive skills by helping them build the auditory map of the target language for their vocal tract practicing, starting with mapping of the L1 phonological inventory.

One of the complicated parts to implementing the virtual talking head framework into classroom environment is the necessity for teacher training. Teachers need to be educated in the area of acoustic and articulatory phonetics in order to skillfully approach learner training. It is then teacher's responsibility to combine their knowledge of the articulatory-acoustic relations to successfully guide learners during the acquisition of the appropriate articulatory gestures. The teacher can experiment with the virtual talking head in order to find the most successful facilitating strategies (Badin et al., 1998). It would be useful for instructors to include diagnostic testing of the learner target group to better understand their needs, and find room for improvement. Then, the classroom can be oriented to address specific needs of the learner group, and teacher can set reachable goals for a set amount of time to ensure productiveness. Even though such setup would require more time and resources, it would ensure good quality pronunciation practice for the learners, and a fruitful research environment for the teacher. This would facilitate more knowledge gain of how virtual talking heads function, in order to develop strategies with supporting evidence to guarantee future learner advancement. Developing a widely-available web-implemented interface would also be very beneficial in the realm of pronunciation learning (I-Chen Lin et al., 1999).

The virtual talking head offers a valuable input of audiovisual nature which, apart from facilitating motivation, helps provide for a multitude of learner types. Badin (2008)

claims that the flexibility in the features and capabilities of such model can lead to promising applications in the domain of speech therapy for speech impaired children, perception and production rehabilitation of hearing impaired children, and pronunciation training for second language learners.

CONCLUSIONS

The virtual talking head framework seems to have the characteristics necessary to become a part of essential instructional material in L2 pronunciation learning. There are many positive examples in research that show such visualization has significant contribution within the teaching of pronunciation. The virtual talking head as a pronunciation assistant, provides the practice in working memory structure via both visual and verbal channels, which minimizes the issues of limited capacity (Ahmah Zamzuri, 2013). The model presented is partially inconvenient, due to its requirement for real-time data. That makes it a time-consuming practice which requires more resources and teacher training. Nevertheless, the virtual talking head has a multitude of benefits. It could help instructors evaluate and improve learner's ability to perceive the sounds of the target language. Learners could identify and visualize their own pronunciation difficulties and, at the same time, improve the relationship between speech production and speech perception. The three-dimensional platform of the virtual talking head could result in the necessary positive impact on second language acquisition of pronunciation.

REFERENCES

Abry, C. & Badin, P. (1996). Speech mapping as a framework for an integrated approach to the sensory-motor foundation of language. *1ˢᵗ ETRW on Speech Production Modeling*, May 1996, Autrans, 175-184.

Ahmah Zamzuri, M. A., & Segaran, K. (2013). 3D Talking-Head Mobile App: A conceptual framework for English pronunciation learning among non-native speakers. *English Language Teaching, 6*(8), 66-76.

Badin, P., Bailly G., & Boë L. J. (1998). Towards the use of a Virtual Talking Head and of Speech Mapping tools for pronunciation training, *Proceedings of the ESCA Tutorial and Research Workshop on Speech Technology in Language Learning, 98*, 167-170.

Badin, P., Elisei, F., Bailly, G., & Tarabalka, Y. (2008). An audiovisual talking head foraugmented speech generation: Models and animations based on a real speaker'sarticulatory data. In *Articulated Motion and Deformable Objects*, Proceedings, 5098, 132–143.

Engwall, O. (2003). Combining MRI, EMA & EPG measurements in a three-dimensional tongue model. *Speech Communication, 41*, 303-329.

Lin, I. C., Hung, C. S., Yang, T. J., & Ming, O. Y. (1999). A speech driven talking head system based on a single face image. *Computer Graphics and Applications, Proceedings*. Seventh Pacific Conference Seoul, 43-49.

Skoyles, J. R. (2010). Mapping of heard speech into articulation information and speech acquisition. *Proceedings of the National Academy of Sciences of the United States of America, 107*(18), E73. http://doi.org/10.1073/pnas.1003007107

Zhu, Y., Fung, A. S. L., & Wang, H. (2012). Memorization effects of pronunciation and strokeorder animation in digital flashcards. *CALICO Journal, 29*(3), 563–577. Retrieved from http://www.jstor.org/stable/calicojournal.29.3.563

Nguyen, P. (2016). eEnglish by Pronunciation Power [Review]. 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. 290-295). Ames, IA: Iowa State University.

SOFTWARE REVIEW

eEnglish by Pronunciation Power
Phuong Nguyen, Iowa State University

Increasing interest in pronunciation teaching and the growth of computer technology has facilitated the development of various software and websites focusing on pronunciation. In fact, teachers and learners of English as a second language (ESL) have been presented with a proliferation of pronunciation software such as *Accent Master, American Speechsounds, Better Accent Tutor, Connected Speech, Pronunciation Power*, to name but a few. A recent website, *eEnglish.com*, published by *Pronunciation Power*, an established standalone pronunciation practice program, offers an inexpensive and engaging experience to independent ESL learners who wish to improve their pronunciation.

DESCRIPTION

For USD 8.99 a month, learners can have access to various packages for pronunciation, vocabulary and grammar, namely, *Pronunciation Power 1* (PP1) and *Pronunciation Power 2* (PP2), *Pronunciation Power (PP) Idioms*, *8 in 1 Dictionary*, *Beginner Grammar with Color Key*, *Vocabulary Builder*, *Speech Test*, and *Phonics Games*. Prior to their first pronunciation lesson, learners are required to take a speech test to identify the sounds that are more challenging to them so that they can focus on these sounds while practicing. In this test, learners are asked to read and record 52 sentences containing English sounds, including 18 vowels, 23 consonants, and 11 cluster sounds. The recordings are then analyzed with an automatic speech recognizer (ASR). Learners are then presented with a list of all English sounds in phonetic symbols, each underlined with a different color indicating whether this sound can be ignored, needs more practice, or needs additional attention. After that, learners can proceed to other packages to improve their pronunciation, vocabulary, and grammar.

Two packages specifically designed for pronunciation are PP1 for beginner and intermediate levels and PP2 for intermediate and advanced learners. In these packages, learners are provided with lessons on each individual sound as well as many exercises for practice on the sound. In the *Lessons* mode, learners can see illustrations of how the sound is produced from both front and side view while they listen to the sound and then access to the *Speech Analysis* function, which gives them a spectrogram of the target sound. At this point, learners can record their sound, which is then analyzed by the ASR, and compare their spectrogram to the virtual instructor's (Figure 1). In the *Exercises* mode, learners can listen to sample words and comparative words of the sound they are practicing and may record them to compare with the words produced by the virtual instructor. The listening discriminating exercises give learners some minimal pair practice in which they listen to two individual words containing the sound and complete a sentence with a

290

word they hear. Additionally, learners can practice the target sound at suprasegmental levels, either sentence stress, word linking, intonation and rhythm, in S.T.A.I.R. (Stress, Timing, Articulation, Intonation and pitch, and Rhythm) exercises. An example of a practice in intonation is presented in Figure 2. The final type of exercise is sentence practice in which learners listen to ten sentences containing the target sound, make recordings of these sentences and compare them to the model sentences.
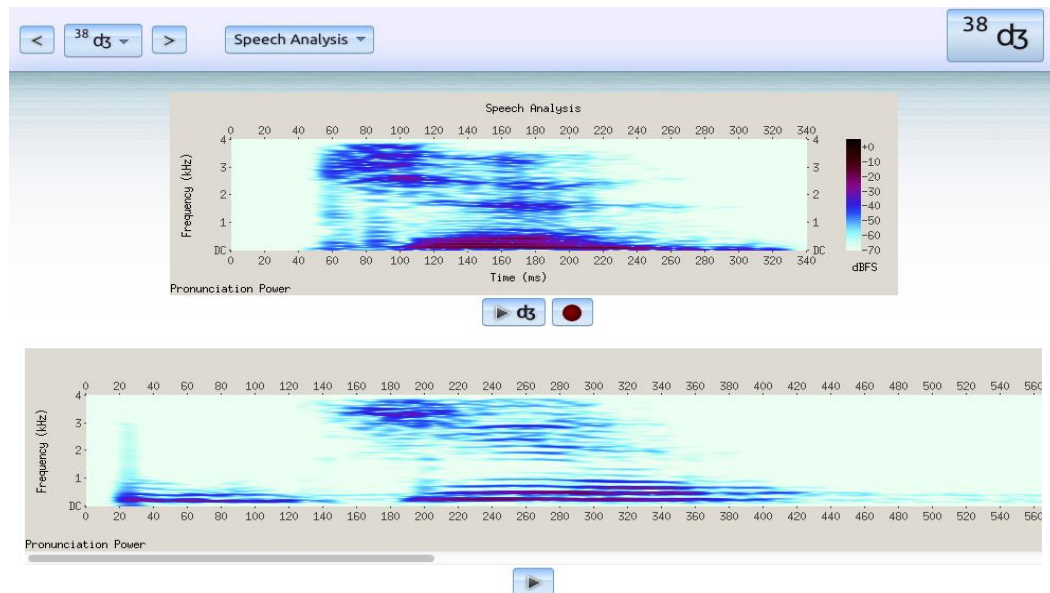

Figure 1. Spectrogram comparing model pronunciation (above) with learners' (below)
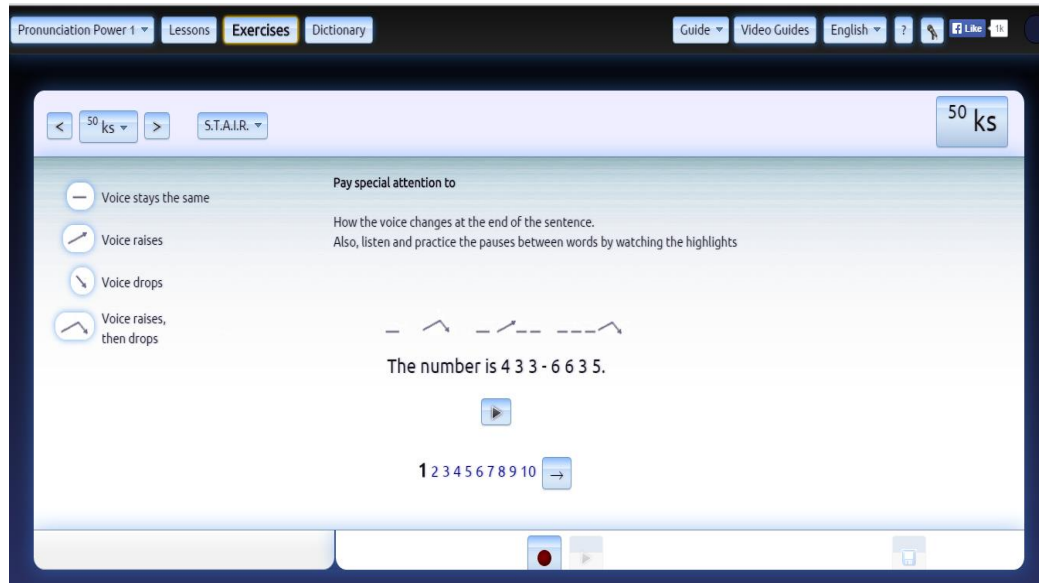
Figure 2. An example practice in S.T.A.I.R

Other packages provided by *eEnglish* include *PP Idioms* with lessons for 104 idioms, each of which is accompanied by a visual illustration, an audio - written example, and many multiple choice exercises which can be scored automatically and provide learners with feedback on their incorrect answers. The *8 in 1 Dictionary* enables learners to search the meaning of a word alphabetically or phonetically and provides them with an explanation in an addition of 12 languages, mostly with a visual illustration (Figure 3). In addition, a different site on 20 basic English grammatical features, accompanied by a pretest, many exercises and a post test, is linked to *eEnglish* to help beginner learners improve their grammar. To improve their vocabulary and to assess their master of the English sounds, learners can have access to *Vocabulary Builder* which provides translations of English words to 12 languages and allows them to make recordings of individual words which will be analyzed by the ASR to identify sounds problematic to them. The last feature is *Phonics Games* in which learners are presented with words visually, logographically and aurally and asked to identify the sound it contains.
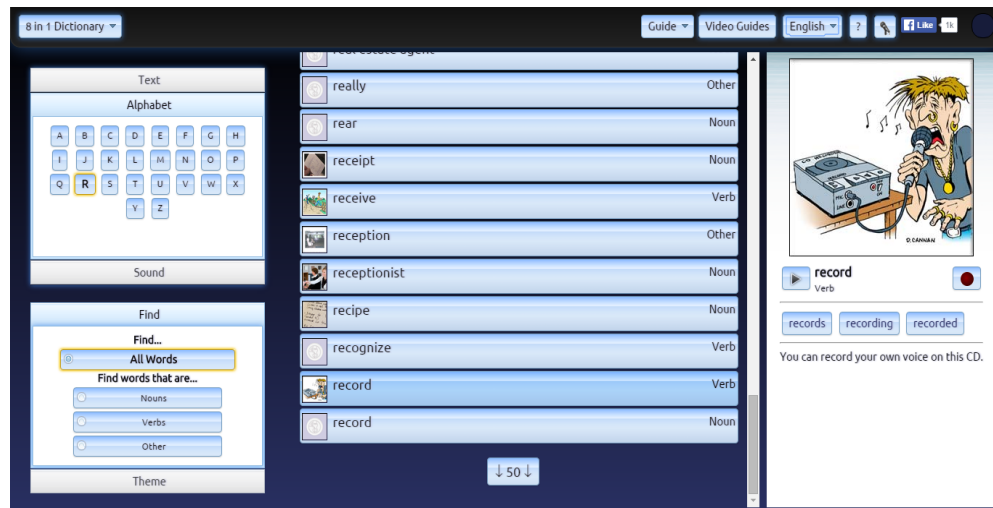
Figure 3. A screenshot of *8 in 1 Dictionary*

STRENGTHS

One of the most useful features of this online learning site is the speech test administered at the beginning of the course. In this test, learners are diagnosed with problematic sounds which they can focus on while working on the site by producing multiple sentences containing the English sounds, a method supported by pronunciation researchers (Levis, 2007). Levis (2007) has argued that computer-assisted pronunciation teaching (CAPT) system should "assist learners and teachers in prioritizing pronunciation topics by channeling learners toward typical vowel and consonant errors for their language backgrounds" (p. 188). This diagnostic test could provide more personalized pronunciation lessons and motivates learners to choose the content that meets their needs.

Another advantage of this site is that, since it incorporates *Pronunciation Power*, it not only focuses on the segmental level but also provides practice at suprasegmental features such as sentence stress, intonation, rhythm and word linking, as evident in exercises in S.T.A.I.R. These features are presented visually with icons such as arrows showing rising or falling intonation, or big and smaller dots showing stress, which have been proved to be valuable pedagogically (Chun, 2007). Although the exercises, focusing on only one suprasegmental feature for each target sound, are not communicative, this inclusion is in line with suggestions made by many pronunciation researchers that suprasegmental features should also be attended to in pronunciation instruction (Celce-Murcia, Brinton, Goodwin, & Griner, 2010; Derwing & Rossiter, 2002).

In addition, a strong characteristic of *eEnglish* is the logical order of a lesson's components. For example, learners are first instructed how the target sound is produced and then presented with listening exercises to help them distinguish contrastive minimal pairs. In these stages, learners can repeat what they hear to practice producing the sounds either in words or sentences and

293

replay the recordings to compare their pronunciation with that of the native speaker instructor. Although more communicative activities in the production stages are not included, the lessons are logically organized from description and analysis, listening discrimination and controlled practice of sounds following the teaching phases suggested by pronunciation researchers (Celce-Murcia et al., 2010).

The ability to give instant feedback is another strength of this online learning site as supported by pronunciation searchers (Neri, Cucchiarini, Strik, & Boves, 2002). Since Flege and Wang (1989) argued that acquiring pronunciation, similar to acquiring syntax in that students need help noticing what they are doing, giving instant feedback can facilitate students' ability to identify through self-monitoring (Celce-Murcia et al., 2010).

Last but not least, in inclusion of various features in the site might motivate learners and facilitate independent learning. Learners can have access not only to pronunciation lessons and practice but also to grammar, vocabulary, dictionary and relatively engaging pronunciation games, which hopefully help them gain other aspects of the language necessary to speaking. The information about their usage, test results and guide achievements can help students keep track of and set goals for their learning.

POSSIBLE IMPROVEMENTS

Some improvements that the site's developer can make are related to the type of feedback to learners' production, the attractiveness and usefulness of the games, the accuracy of instructions in languages other than English and the compatibility of speech analysis software.

First and most importantly, the feedback given to learners could be more comprehensible and thorough to learners, which is usually a problem for CAPT systems (Levis, 2007). First, learners are presented with feedback only to their speech test, their production of the target sounds in isolation and the listening discrimination activities. Second, when learners' production of a single sound is analyzed, a spectrogram of their sound is given, which is usually neither transparently interpretable to learners (Levis, 2007; Neri et al., 2002) nor accurate (Kim, 2006), Also, it is difficult, even for native speakers of English, to achieve the same spectrogram as the model speaker. Therefore, spectrogram feedback is of little value to learners. Third, although it is good that learners have opportunities to record their production and compare it to the model speaker's, it would be more beneficial if an accurate ASR system was incorporated to check for comprehensibility, since sometimes students fail to distinguish the difference in the sounds they produce and more appropriate ones.

Second, the games that learners have access to could be more motivating and useful. Although the pronunciation games (*Phonic games*) can be fun and engaging, they focus only on learners' receptive skills. Incorporating games requiring learners to produce the target sounds in a various

294

game formats would be more attractive and useful to learners, since they will have opportunities to produce the sounds they have just learnt and check if their learning is successful.

Other necessary improvements include the accuracy of instructions in other languages and the compatibility of speech analysis software. It is useful to have instructions in PP1 and PP2 in 12 different languages. However, the translation of these instructions, at least to Vietnamese, is incomplete and inaccurate, which can be confusing, especially to learners of lower proficiency levels. Also, *eEnglish*'s developers would have to work more to improve the compatibility of the speech analysis software so that it can work with Window 7 or Vista, 64 bits.

CONCLUSION

Despite the aforementioned drawbacks, *eEnglish* can be considered a useful tool for improving pronunciation, considering its reasonable price and useful features. While prices of other pronunciation programs or the standalone *Pronunciation Powers* are more pricy, starting from approximately USD 75 (*American Speechsounds*), *eEnglish* offers learners access to many packages at a less expensive price. The diagnostic speech test, the instant feedback provided by ASR system and the integration of multiple features other than pronunciation are some highlights of the website. This website is appropriate for independent learners who wish to practice pronunciation at the time convenient to them.

REFERENCES

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

Chun, D. M. (2007). Come ride the wave: But where is it taking us? *CALICO, 24*(2), 239-252. doi:10.1558/cj.v24i2.239-252

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.

Kim, I. S. (2006). Automatic speech recognition reliability and pedagogical implications for teaching pronunciation. *Educational Technology & Society, 9*(1), 322-334.

Levis, J. (2007). Computer Technology in Teaching and Researching Pronunciation. *Annual Review of Applied Linguistics, 27*. doi:10.1017/s0267190508070098

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. doi:10.1076/call.15.5.441.13473

BOOK REVIEW

Derwing, M. T., & Munro, J. M. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. Philadelphia, PA: John Benjamins Publishing Company. 208 pages. ISBN 978-9027213273

Sonca Vo, Iowa State University

Achieving a native-like accent is not easy for many non-native English adult learners (Levis & Moyer, 2014), and second language (L2) speech produced by adult learners is usually characterized with incorrect pronunciation of segmentals (e.g., consonants, vowels) (Flege, Munro, & MacKay, 1995) and suprasegmentals (e.g., stress) (Derwing, Munro, Foote, Waugh, & Fleming, 2014) of the L2 phonological system. Therefore, on the way to helping L2 pronunciation researchers and teachers understand the *whys* and *hows* of assisting L2 learners to develop their oral communication more effectively, Tracey Derwing and Murray Munro has introduced the book *Pronunciation Fundamentals: Evidence-based Perspectives for L2 Teaching and Research*. The book is in a series edited by Nina Spada and Nelleke Van Deusen-Scholl and published by John Benjamins publishing company in Amsterdam/Philadelphia in 2015. It is organized into ten chapters that address the key components of intelligibility and comprehensibility, presented mainly with theory and research in English as second language pronunciation with the strengths and weaknesses of specific studies in the view of the Intelligibility Principle (Levis, 2005), all of which ultimately inform pronunciation instruction of not only English but other languages as well.

The first chapter of the book "Key Concepts" is short but important in providing key terms in L2 pronunciation such as segmentals, suprasegmentals, comprehensibility, fluency, foreign accents, accentedness, nativeness and intelligibility principles, minimal pair practice, phonetics, acoustic measurement, and functional load. Then, to contextualize the current practice of pronunciation teaching and research, Chapter 2, "Historical Overview of Pronunciation", surveys the historical developments of English pronunciation teaching and presents early studies that are neglected but still crucial in shaping the current affairs in the field of pronunciation teaching and research. Four main themes that inform pronunciation instruction include descriptions of the English sound system in terms of orthography and sociolinguistic influences; classroom pronunciation teaching materials and technological developments in pronunciation teaching; phonological theories in L2 speech perception and production; and empirical studies on pronunciation instruction.

In the third chapter, "A Pedagogical Perspective on L2 Phonetic Acquisition", the authors provide insights into important aspects of the L2 phonological acquisition process by discussing the role of age; motivational influences; aptitude for pronunciation; and instruction, all of which are valuable for pronunciation researchers and instructors to develop curricular, choose activities,

and provide feedback to learners. This chapter emphasizes the importance of Intelligibility Principle over the Nativeness Principle in communicative pronunciation instruction, especially for adult pronunciation learning because late learners rarely become native-like (Levis & Moyer, 2014). Therefore, native-like acquisition should not be a focus in pronunciation instruction, but intelligible speech should be targeted and could be achievable if learners are given more opportunities to interact regularly in their L2.

In Chapter 4, "Pronunciation Errors and Error Gravity", the authors explore L2 pronunciation errors that interfere with L2 intelligibility such as segmental errors, prosodic errors, aspects of fluency, voice quality, and several non-linguistic and paralinguistic aspects of L2 speech. Contrastive Analysis Hypothesis (CAH) and Error Analysis (EA) are presented together with prior studies to provide information about first language (L1) effect on L2 speech. The chapter specifically discusses the weaknesses of CAH and EA because "perception and production are not considered separately in these approaches", and CAH and EA offer no obvious way of "accounting for changes in learner performance over time"; "individual learner differences" (p. 65), and "underlying cognitive process that leads to pronunciation errors" (p. 66). The authors mention that apart from the L1 interference, individual learner variability is also an influential factor. Therefore, in order to address pronunciation errors "instructors should give considerable weight to individual differences and to variable error gravity" (p. 76) to help learners increase their L2 comprehensibility and intelligibility.

Chapter 5, "Pronunciation Instruction Research", presents surveys of pronunciation features that teachers prioritize during their pronunciation instruction and whether teachers have sufficient training to teach pronunciation. Another interesting question raised in this chapter is whether non-native speakers should teach L2 pronunciation. The authors point out that the assumption that "only a native speaker has the wherewithal to effectively teach pronunciation" is "faulty" (p. 81) because "many so-called NSs can be far less intelligible in global settings than well-educated proficient speakers of a second language" (Moussu & Llurda, 2008, p. 318). A large section of the chapter covers classroom-based intervention studies, showing that instruction leads to a significant improvement in L2 speakers' perception and production.

Chapter 6, "Assessment of L2 Pronunciation", discusses the role of pronunciation in language assessment, pronunciation assessment instruments in high- and low-stakes testing situations, and important issues of validity and reliability in pronunciation assessment. For classroom instruction, the authors suggest that teachers adapt published resources and use them for needs assessments, formative assessments, and summative assessments according to different language learners and teaching situations. The chapter also emphasizes the increasing importance and reliability of automatic assessment in evaluating test-takers' oral performances in high-stakes testing. However, the complexities in evaluating speech samples that just focuses on two aspects of intelligibility and comprehensibility lead to the fact that 'pronunciation testing remains an underdeveloped aspect of language assessment' (p. 119).

The next topic discussed in Chapter 7 "Technology in L2 Pronunciation Instruction" is the role of technology in teaching and assessing L2 pronunciation. In line with what Chapelle (2003) argues, which is TESOL educators 'need to be critically aware of the connections among technology, culture, and ideology, and specifically about the ways in which technology amplifies and constrains aspects of language learning and research' (p. 9), the authors recommend that teachers have to understand the foundations of pronunciation research and have good pedagogical knowledge to fully take advantage of the strengths that technology offers to enhance pronunciation learning and teaching.

Chapter 8, "Social Aspects of Accent", discusses issues such as social impact of speaking with a foreign accent, listeners' attitudes towards accented speech, L2 accented speech and speaker identity, accent discrimination in the workplace and the field of language teaching, the training of native English speakers to become better listeners, and the Willingness to Communicate framework applied to both native and non-native speakers. From all of these issues, implications for instruction are suggested. Ethical considerations in L2 pronunciation instruction in accent reduction programs and misinformed practices in pronunciation instruction are presented in Chapter 9 "The Ethics of Second Language Accent Reduction". Considering the complexities in pronunciation instruction, the authors recommend that TESL programs should well equip their future language teaching professionals with better understanding of the principles of pronunciation teaching and of the difference between accent and intelligibility.

Chapter 10, "Future Directions", concludes the book with recommendations for future L2 pronunciation research regarding factors (e.g., linguistic aspects of L2 speech) affecting L2 intelligibility and comprehensibility; listeners' processing abilities; and situation-related understanding issues. Research on pronunciation teaching strategies and techniques and longitudinal studies as well as directions for teaching with a more focus on the Intelligibility Principle are suggested. The book ends with directions for assessment, technology, and the larger society with an argument that 'all human communication is a two-way street; all interlocutors share responsibility for the outcome of any exchange' (p. 172). Therefore, successful communication should be a joint effort from learners, native speakers, instructors, and the public.

With the culmination of L2 phonology theories that date back to the 1980s and are the foundations for early as well as recent empirical studies on L2 pronunciation, the book is intended for scholars and professionals who are dealing with L2 learners and want to know more about the phonological development of L2 learners' interlanguage, thereby leading to better understanding of pronunciation errors in L2 speech. Furthermore, the book does offer insightful contributions to L2 pronunciation instruction by providing a detailed synthesis of empirical research studies in L2 pronunciation instruction over the past decades. The discussions of pronunciation research and approachable teaching techniques provide a crucial source for novice as well as experienced pronunciation teachers to better understand important aspects of pronunciation that they should prioritize during their instruction.

The book also does an excellent job of making sure key vocabulary is defined and contextualized in both early and recent L2 pronunciation research studies. Thus, readers greatly benefit from those early but very important studies because they shape readers' understanding of current affairs in L2 pronunciation research and teaching. Last but not least, written in accessible language while presenting complex information about L2 phonology theories and pronunciation research, the book can be therefore perfectly used as a textbook in a course on pronunciation teaching. However, as the authors have indicated that ideas for pronunciation activities and information on the International Phonetic Alphabet (IPA) are not available to be referenced from the book, instructors need to supplement the book with those resources in order to fully assist students without a basic knowledge of linguistics, phonetics and the IPA.

Overall, written by the two scholars who have been devoted to L2 pronunciation teaching and research for many years, the book is an excellent source for pronunciation researchers, ESL teachers in the field of L2 pronunciation, or students in TESL/applied linguistics programs who would like to comprehensively review the issues of history, pedagogy, linguistics, and social and ethical aspects in teaching and assessing L2 pronunciation. Linking different phonological theories and empirical studies in L2 pronunciation research and teaching, especially with an inclusion of pedagogical implications at the end of each chapter, the book successfully shows how important pronunciation is and how pronunciation could be taught effectively.

REFERENCES

Chapelle, C. A. (2003). *English language learning and technology: Lectures on applied linguistics in the age of information and communication technology*. Amsterdam: John Benjamins.

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

Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995). The effect of age of second language learning on the production of English consonants. *Speech Communication, 16,* 1-26.

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

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

Moussu, L., & Llurda, E. (2008). Non-native English-speaking English language teachers: History and research. *Language Teaching, 41*(3), 315-348.

SOFTWARE REVIEW

Reading Assistant^{TM}
Yongkook Won, Iowa State University

It is not always the case of medical science where the side effect of a product becomes a new innovation with an effect that was not intended to be developed. The online English learning program *Reading Assistant*^{TM}, developed by Scientific Learning Corporation, is advertised to help native speakers of English learners to acquire phonics skills, which help to learn correspondence between sounds and the spelling patterns, and eventually to improve oral reading ability (Beattie & Chevalier, 2012). However, the product is also reported to be used in an English as a foreign language (EFL) context to help non-native speakers of English learners to acquire pronunciation of English words and to improve English pronunciation and speaking (Jeong, 2010). Those EFL learners use phonics method not only for oral reading ability improvement, but also for pronunciation practice (Kim, 2005; Kuo, 2011). Even though *Reading Assistant*^{TM} is not specifically developed for non-native speakers of English learners, it is reported that there are increasing numbers of English learners who use the product to improve their oral language skills (Neuro Science Learning, 2015).

Reading Assistant^{TM} consists of three main components: reading materials, automatic English speech recognition system, and the reading fluency review with pronunciation error records. First, the online program has four levels of reading materials, which are modified from famous novels, news magazines, and other materials with familiar topics to learners; thus, the program seems to be suitable to anyone from elementary school students to adult English learners. Second, an English speech recognition system, which uses PocketSphinx speech recognizer (Walker et al, 2004), is applied to *Reading Assistant*^{TM} and it helps decide whether the readers, or the language learners, pronounce the given words appropriately. This speech recognition engine phonetically compares the learner's reading pronunciation against a pronunciation dictionary in the program. As illustrated in Figure 1 and Figure 2, students can record and listen to their oral readings and get feedback from the evaluation engine in *Reading Assistant*^{TM}. Finally, there is a review session that provides feedback to the students who read aloud given texts in the program. When the reading activities are done, every word of which learners mispronounced or had difficulties in pronouncing are presented with different diagnostic symbols depending on the types of errors. The overall study flow of *Reading Assistant*^{TM} is shown in Figure 2.
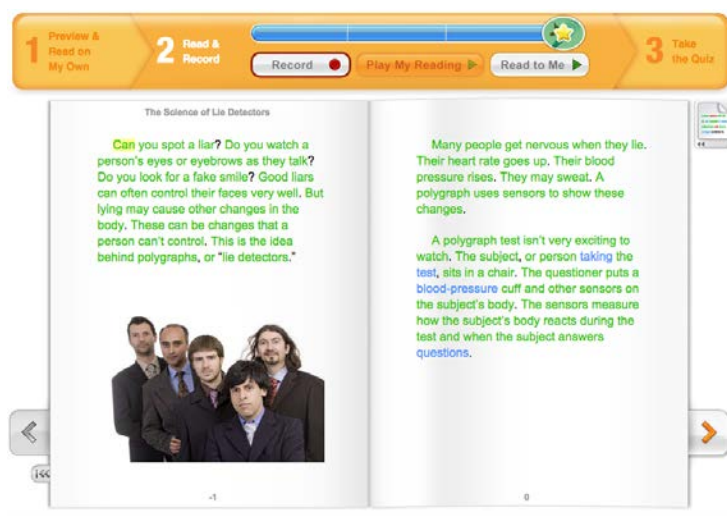
Figure 1. *Reading Assistant*™ software, read & record stage screenshot. Words in blue fonts are the expressions that students mispronounced.
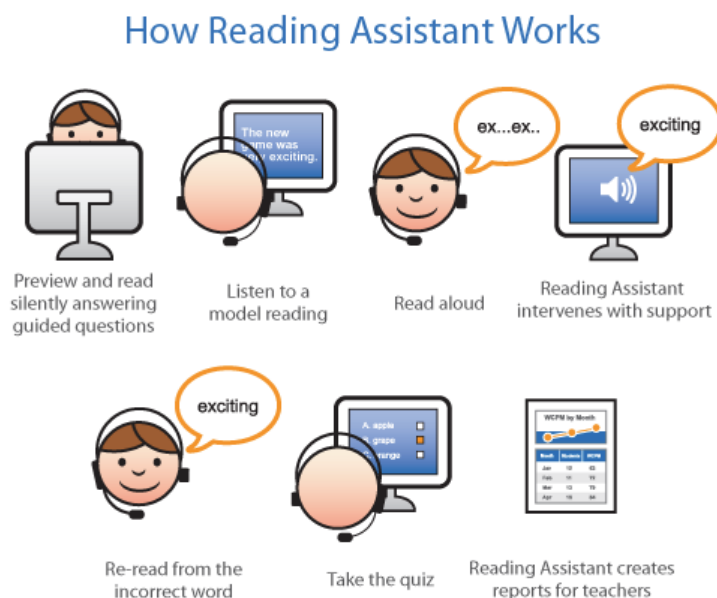

Figure 2. *Reading Assistant*™ software, study flow
(http://www.scilearn.com/products/reading-assistant)

Each component of *Reading Assistant*™ seems to have the following three main merits for EFL learners. First, the reading materials provided in the program help EFL learners not to lose their motivation when they are doing pronunciation practice. As teachers in an EFL context are reluctant to give regular pronunciation instruction (Derwing & Munro,

2015), pronunciation practices in EFL context are usually done by the learners themselves with drill-based pronunciation learning materials. These repeated drill exercises without any assistance could make learners easily feel exhausted and lose their motivation to continue practicing. Thus, the reading materials with audio files in *Reading Assistant*™ could be complementary to the drill-based language training. Furthermore, the reading contents in *Reading Assistant*™ simply can be replaced or customized with other reading materials when requested by other language learning materials developers or users, thus making the reading contents provided to learners without limits. This adaptation of content materials seems to be one of the merits of this type of module-based program. To increase the flexibility of the materials, however, the chances of applying speech recognition systems to supra-segmental level seems to be decreased, because the currently employed dictionary-based pronunciation rules are better for word-level than for sentence-level speech recognition.

Second, the English speech recognition system in *Reading Assistant*™ are reported to work well to recognize the pronunciation of English learners as well as that of English native speakers and to catch the mispronunciation of both speaker groups (Beattie & Chevalier, 2012). The speech recognition engine does not evaluate the proficiency level of the reader's pronunciation, and it only needs to decide whether the pronunciation is acceptable or not. The reported false negative error rates, which give warnings when there are no real pronunciation errors, are less than 3% when tested with audio files of middle school students including both English native speakers and English language learning students, and around 1% with the audio files of children and adults of native English speakers in the United States (Beattie & Chevalier, 2012). For this reason, even though *Reading Assistant*™ was originally developed for native speakers of English who struggle with reading texts, its speech recognition engine seems to be used in judging the pronunciation quality of English language learners as well. Being said that, readers should be cautioned not to consider low false negative rates as the accuracy of the speech recognition system. There could be still higher possibilities of not providing feedback to students, or false positives, when they make pronunciation errors.

Finally, the last beneficial function of *Reading Assistant*™ is the review of students' pronunciation accuracy. The ASR system not only provides immediate feedback when learners mispronounce the given words in the reading texts, but also gives a summary page of the correct and incorrect pronunciation with different color fonts for each error type (see Figure 1 for the mispronounced words in the given context and see Figure 3 for the summary of overall performance). The fluency report page provides the overall picture of the learner's pronunciation patterns by looking at what types of words were mispronounced and guides learners where to put more attention. Even though the online program is developed to provide self-directed learning, human assistance with the summary page would increase the effectiveness of the training. Because young learners, or novice learners, cannot easily find the patterns of their pronunciation errors, it would be more efficient to have pronunciation tutoring together with self-practice of the program.
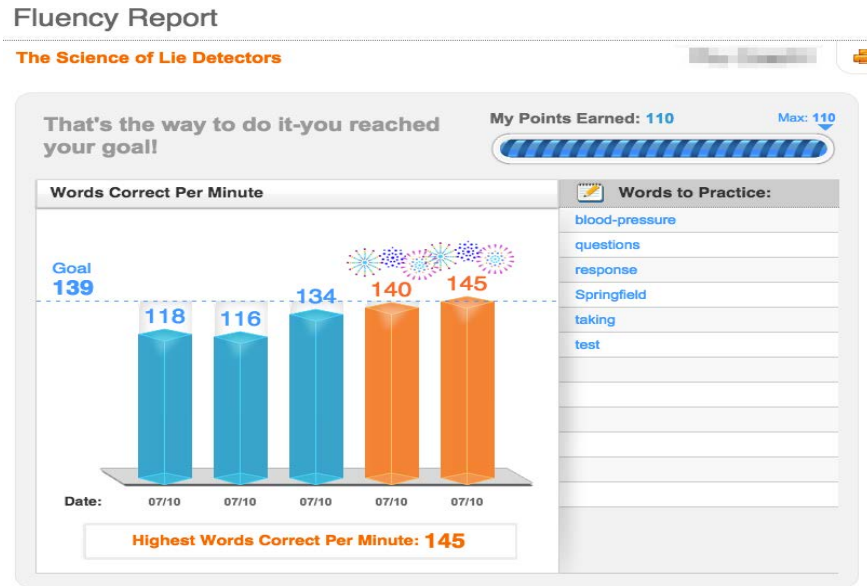
Figure 3. *Reading Assistant*™ software, fluency report screenshot

As is often the case with computer assisted language learning (CALL) programs, one overarching criticism of *Reading Assistant*™ is that there is less interaction in the language learning and learners do not have immediate help from human teachers. Although the program provides immediate feedback for every single mispronunciation, it may not be as adaptive as human teachers in adjusting the contents to learners' current speaking status and providing an adequate level of study materials. In addition, the pronunciation in spoken communication may not be practiced with *Reading Assistant*™, because the program only provides pronunciation practice with reading texts and pronouncing words by reading them are different from the pronunciation in an oral communication (Swerts et al., 1996). In addition, *Reading Assistant*™ may not provide enough input to the learners who generally needs huge amount of aural inputs before they produce oral outputs, because the English native speaker's pronunciation is only given when learners make pronunciation mistakes or choose the text listening option.

Despite the aforementioned limitations, however, this automated reading assistant program seems to make a valuable contribution to the field of speaking as well as reading education for EFL learners in that it provides opportunities to the learners to autonomously study spoken language which was usually practiced only with the help of human teachers or tutors. This individualized learning environment is believed to reduce the anxiety level of learners which they usually have when they are in public or in front of other human beings, such as teachers or friends. In addition, the immediate feedback of the mispronunciation may reduce the burden of looking up dictionaries when students encounter unfamiliar words. This reading program may provide good chances to improve pronunciation quality of the EFL learners, especially of those who are in an EFL context and who generally learn written English before spoken English.

REFERENCES

Beattie, V., & Chevalier, S. (2012). *Reading Assistant: Technology for guided oral reading*. Scientific Learning.

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

Jeong, J. (2010, June 25). Neuroscience Learning English oral reading program conference. *The Korea Economic Daily*. Retrieved from http://www.hankyung.com/news/app/newsview.php?type=2&aid=201006258404i &nid=910&sid=0104

Kim, B.-s. (2005). A study on the effective phonics strategies. *English Language Education*, *31*, 105-123.

Kuo, L.-C. (2011). The role and efficacy of phonics instruction in the early literacy development of young Taiwanese EFL learners (Doctoral dissertation, University of Warwick). Retrieved from http://wrap.warwick.ac.uk/45408/1/WRAP_THESIS_Kuo_2011.pdf

Neuro Science Learning (2015). *Help them amaze you: Neuro Science Learning*. Retrieved from http://www.nslearning.co.kr/NSL-company.pdf

Swerts, M., Strangert, E., & Heldner, M. (1996). *F_0 declination in read-aloud and spontaneous speech.* Paper presented at the ICSLP 96.

Walker, W., Lamere, P., Kwok, P., Raj, B., Singh, R., Gouvea, E., . . . Woelfel, J. (2004). *Sphinx-4: A flexible open source framework for speech recognition*. Sun Microsystems.

Zhou, Z. (2016). Speaking section in English speaking and writing test - ESWT [Review]. 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. 305-309). Ames, IA: Iowa State University.

SOFTWARE REVIEW

Speaking Section in English Speaking and Writing Test (ESWT)
Ziwei Zhou, Iowa State University

INTRODUCTION

The English Speaking and Writing Test (ESWT) is a web-based, free of charge, and low-stake testing program, developed in Pai Chai University in Daejeon, South Korea to provide appropriate assessment of local university students' speaking and writing proficiency (Kim, 2011). Its development was originally motivated by the various advantages computer-assisted language testing (CALT) offers such as assessing large number of test takers and the ability to track students' improvement. Another motivation derived from the observation that the expensive test fees in many popular commercial testing products caused financial burdens for students (Kim, 2011). Therefore, a test that is sensitive to local context is needed to assess students speaking and writing proficiency on the one hand, and provide constructive feedback to track and facilitate students' English language development on the other (Kim, 2011).

SPEAKING TASKS OF THE ESWT

The speaking consists of 4 tasks that prompt test takers to produce extended responses with open-ended questions. In the first task, test-takers are asked to introduce themselves in 45 seconds, with 10 seconds preparation time. When they are ready to respond, they need to hit the RECORD button. When they finish recording, they hit the STOP button. They are asked to speak at least 30 seconds. While they are responding, they can see their recording volume as well as remaining time on the screen. In the second task, test-takers need to narrate a story based on a set of six pictures. They have 30 seconds preparation time and 60 seconds response time. The pictures involve common topics for university students. The third task provides test-takers with visuals such as tables, bar graphs, etc. and requires them to describe the visuals. In the last task, test-takers are prompted to give their opinions on familiar topics that are closely associated with their personal life (e.g. "What is your major? How will you contribute to the society with your major? Why?") or some contentious issues that are common to the them (e.g. political issues between North and South Korea). They have 30 seconds to plan their answers and 60 seconds for answering. They are also required to speak at least 30 seconds.

VALIDATION

The development of the ESWT adopted Chapelle, Jamieson, and Hegelheimer's (2003) suggestion to collect validity argument evidence during the entire process of test development. It integrated Davidson and Lynch's (2002) test-spec approach, ADDIE model (i.e. Analyze, Design, Development, Implementation, and Evaluation), and Bachman and Palmer's (1996) Test Usefulness framework (Kim, 2011).

According to Kim (2011), the constructs of the speaking section, including fluency, functional competence, pronunciation, grammar, vocabulary and expressions, and coherence, were based on ACTFL proficiency guidelines in speaking. Tasks were designed on the basis of the specified constructs in the scoring rubric. Additionally, correspondence in terms of topics, situations, sources, preparation and response time, and answering methods between the test tasks and tasks in the TLU domain were attended to so as to strengthen context validity (Kim 2006b; Kim, 2011). The main test was implemented via multimedia authoring tool and administered by FTP-based management system. Finally, the test was evaluated by statistical analysis of test scores as well as feedback from students.

EVALUATION

Construct Validity

The seminal work of Cronbach and Meehle (1955) defined construct as "some postulated attributes of people assumed to be reflected in test performance" (p. 178). Therefore, the construct model concerns with indirect measures of abilities or attributes of human behavior. In his adoption of the Usefulness Analysis Table (Bachman & Palmer, 1996) to evaluate the EWST, Kim (2011) established construct validity based on task design and difficulty as well as interface design. He pointed out that this quality of usefulness was supported by empirical evidence showing that students perceived test tasks as acceptable and test interface as satisfying. In defining the constructs, Kim (2011) only made brief reference to scale descriptors in ACTFL, TSE, and TWE. Beyond this, no statements were made about test takers' certain attribute as reflected in their test performance. The construct definition in the test development and validation process remained largely absent. The fundamental rationale of the author's position in conceptualizing the constructs was also unstated. This lack of explicit attendance to construct definition may partly explain the lack of reference to any construct theory in Kim (2011).

The sole reference to ACTFL, TSE, and TWE may be problematic because the particular contexts where the ESWT was used may entail different requirements for university students in Korea than the U.S. where these proficiency rubrics and rating scales were developed. Even though the test was used for low-stake purposes, solidifying the theoretical rationales and beliefs by referencing to "theory of the construct" and "construct theories" is also needed since it forms the basis for test specification as well as the hypothesized relations in the nomological network and test constructs and other constructs (Messick, 1989; Chapelle et al., 2003).

Content Validity

According to Luoma (2004), comprehensive content coverage in relation to the definition of test purpose should be the major validity concern of speaking assessment. Such relation can be carefully investigated by delineating task features between the test and non-test situations. In the ESWT, Kim (2011) did not seem to provide sufficient evidence to show that the test tasks were representative of the TLU domain. Though evidence from university teachers' opinions and students' perceptions are crucial, domain description

and modeling has not been reported. Task representativeness should be carefully addressed by systematic and exhaustive attempts to map features between test and non-test situations with reference to refined framework of task characteristics, interactions, responses, and evaluations of tasks. Granted, even though the task feature correspondence is carefully drawn, some task in the TLU domains are impossible to be captured in CALT settings. At the very least, prompts should contain sufficient contextual cues to engage test takers' *discourse domain* (Douglas, 2000) in interacting with the test tasks. Otherwise, there may be no basis to infer that the intended constructs are appropriately and adequately elicited by the test tasks.

Interactiveness and Reliability

Kim (2011) used Bachman and Palmer's Test Usefulness framework to evaluate the development and validation results of EWST (see Figure 1 below).

Quality	Positive Attribute	Negative Attribute
Construct Validity	* The test was developed to elicit only variation by the intended English speaking and writing ability using the spec-driven testing framework. * Task difficulty is acceptable with mean and standard deviation scores of speaking (4.00(.74)) and writing (3.89(.72)). * Test takers' perceptions of task difficulty are acceptable with mean and standard deviation scores of speaking (3.71(.69)) and writing (3.82(.74)). * Test takers are satisfied with the test interface design in general.	* Test takers were a little dissatisfied with the task topics and situations. * Test takers felt that preparation time and writing response time were a little short. * For item #2 and #9, test takers perceived task difficulty contrary to the real test scores.
Reliability	* Test topics and situations were designed to elicit only variation by the intended constructs that were extracted based on construct theories. * Task difficulty in terms of test scores and perceptions was acceptable.	* To elicit precise constructs, constructed-response tasks should have been included.
Authenticity	* The test topics and situations were chosen from appropriate TLU domain. * Test takers felt that tasks were highly relevant to their real life. * Test takers are highly familiar with the Web-based test tool.	* The TLU domain is too wide. * The tasks were not developed to take care of all the English language proficiency of the test takers. * Three test takers failed to respond to the first task.
Interactiveness	None	* There is no interlocutor so that it is impossible to measure the quality.
Impact	* The test offers testing experience to the participants. * The test may achieve positive washback-effects so that they improve their English speaking and writing ability.	* Some students may feel pressure due to the test preparation.
Practicality	* The test was developed by the department faculty with low cost and can be administered at general computer labs without extra cost.	* Rating fees and time are inevitable.

Figure 1. The ESWT usefulness analysis table.

In the ESWT Speaking, quality of interaction was impossible to measure due to the lack of interlocutor (Kim, 2011). Without the interactiveness component, it becomes impossible to investigate the extent to which certain elements in the theoretical construct model is activated during test takers' engagement with the test tasks. In addition, the basis of investigating strategic competence through analyzing test taking processes is utterly discarded. Moreover, even though the test is functioning in the local university setting regularly (J, Kim, personal communication, November 12, 2015), no evidence of performance consistencies is documented. Also, no empirical study results in terms of

rating agreement or reliability coefficient between test and re-test are reported. Kim (2011) established his reliability evidence on the basis of construct relevancy and task difficulty. However, since explicit construct definition and model is lacking, the claim that only relevant and intended constructs are elicited as a function of the choice of test topics and manipulations of test situations seems unconvincing.

CONCLUSION

The ESWT is an up-to-date example of CALT product meeting local demands. Its development and evaluation process incorporated multiple frameworks across disciplines including language assessment, curriculum and instruction, business management, and CALT. The test is authentic since test tasks were chosen from appropriate TLU domain (Kim, 2011). Moreover, the ESWT gives due attention to generating positive impact. Test results are used as materials for the weekly coaching sessions, which facilitates students' self-directed learning (J, Kim, personal communication, November 18, 2015). Finally, the ESWT fulfills its original motivation of development by providing free testing experiences and easy access for all users. Recently, the test developer is collaborating with software engineers to upgrade the program system and implement the test on mobile devices (J, Kim, personal communication, January 20, 2016). For future development, EWST should pay closer attention to construct definition and domain description in order to consolidate and explicate the theoretical rationales and beliefs about the speaking abilities in academic contexts. Such efforts will also help to pinpoint task features at a more refined level, which provides the basis to elicit relevant constructs through test takers' own engagement in the test tasks. Future development and evaluation can also investigate test takers' performance consistencies and rater behaviors across time and settings in order to enhance reliability.

REFERENCES

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice.* Oxford: Oxford University Press.

Chapelle, C. A., Jamieson, J., & Hegelhemer, V. (2003). Validation of a Web-based ESL test. *Language Testing, 20*(4), 409-439.

Davidson, F., & Lynch, B. (2002). *Testcraft: A teacher's guide to writing and using language test specifications.* New Haven, CT and London: Yale University Press.

Douglas, D. (2000). *Assessing language for specific purposes.* Cambridge, UK: Cambridge University Press.

Kim, J. T. (2006a). *The effectiveness of test-takers' participation in development of an innovative web-based speaking test for international teaching assistant in American colleges.* Unpublished dissertation, University of Illinois, Urbana-Champaign.

Kim, J. T. (2006b). Context validity of the Speaking Proficiency English Assessment Kit (SPEAK). *Korean Journal of the Applied Linguistics, 22*(2), 137-158.

Kim, J. T. (2011). The validation process in developing a web-based English speaking and writing test. *Multimedia-Assisted Language Learning, 14*(2), 181-209.

Luoma, S. (2004). *Assessing speaking*. Cambridge, UK: Cambridge University Press.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (pp. 13-103). New York: Macmillan.