

**PROCEEDINGS OF THE 6TH ANNUAL
PRONUNCIATION IN SECOND LANGUAGE
LEARNING AND TEACHING CONFERENCE
(ISSN 2380-9566)**

Editors

John Levis

Rania Mohammed

Manman Qian

Ziwei Zhou

6TH ANNUAL PSLT CONFERENCE

PRONUNCIATION IN SECOND LANGUAGE LEARNING & TEACHING 2014



September 5-6, 2014

1207, 1215, 1217 Education Building University of California,
Santa Barbara



PRONUNCIATION IN SECOND LANGUAGE LEARNING AND TEACHING

6th Annual Proceedings

(Conference held September 2014, Santa Barbara, California)

TABLE OF CONTENTS

A step further.....	1
---------------------	---

John M. Levis, Iowa State University
Rani Mohammed, Iowa State University
Manman Qian, Iowa State University
Ziwei Zhou, Iowa State University

Experimental Studies

Acquisition of English L2 prosody by Italian native speakers: Experimental data and pedagogical implications	15
--	----

Maria Grazia Busà, University of Padova
Antonio Stella, University of Salento

Discourse intonation in L2 academic presentations: A pilot study	27
--	----

Larissa Buss, Concordia University
Walcir Cardoso, Concordia University
Sara Kennedy, Concordia University

Mexican & Chicano Spanish prosody: Differences related to information structure.....	38
--	----

Michael J. Harris, University of California Santa Barbara
Viola G. Miglio, University of California Santa Barbara
Stefan Th. Gries, University of California Santa Barbara

Learner preference and the learning of Japanese rhythm	49
--	----

Naoko Kinoshita, Waseda University

When French becomes tonal: Prosodic transfer from L1 Cantonese and L2 English	63
---	----

Jackson L. Lee, University of Chicago
Stephen Matthews, University of Hong Kong

Are French immersion “accents” unique?	73
--	----

Livia Poljak, Simon Fraser University

Acquisition of word-final devoicing by American learners of Russian	85
---	----

Ala Simonchyk, Indiana University
Isabelle Darcy, Indiana University

Swedish tonal word accents produced by Vietnamese L1 speakers.....	95
Elisabeth Zetterholm, Stockholm University	
Mechtild Tronnier, Lund University	

Technology and L2 Pronunciation

Audiovisual and auditory-only perceptual training: Effects on the pronunciation of French nasal vowels	104
Solène Inceoglu, Rochester Institute of Technology	

Consonant clusters in online L2 teaching: A multilingual approach.....	115
Violeta Martínez-Paricio, Norwegian University of Science and Technology	
Jacques Koreman, Norwegian University of Science and Technology	
Olaf Husby, Norwegian University of Science and Technology	
Jardar Eggesbø Abrahamsen, Norwegian University of Science and Technology	
Øyvind Bech, Norwegian University of Science and Technology	

Automatic speech recognition: Making it work for your pronunciation class	126
Shannon McCrocklin, University of Texas Rio Grand Valley	

Extracting minimal pairs automatically with word frequency and phonetic environment controlled: Introducing a program written in PERL	134
Manman Qian, Iowa State University	

Instructional issues and approaches

A sociocultural view of engagement in the music-based pronunciation classroom.....	143
Catrice Barrett, University of Pennsylvania	

What's hot, what's not: Insights from pronunciation practitioners	154
Donna Brinton, Educational consultant	
Marsha Chan, Sunburst Media and Mission College	

Putting it all together: From pronunciation analysis to pronunciation pedagogy.....	166
Janay Crabtree, University of Virginia	

The role of pitch contours in teaching vowel length distinctions in Japanese	174
Masanori Deguchi, Western Washington University	

A NAIL in the coffin of stress-timed rhythm	184
Wayne B. Dickerson, University of Illinois at Urbana-Champaign	

Advanced adult ESL students' perspectives on the benefits of pronunciation instruction	197
Lynn Henrichsen, Brigham Young University	
Chirstin Stephens, Brigham Young University	

Revisiting the pronunciation of English by speakers from Mainland China.....	206
Meichan Huang, Texas A & M University, Commerce	
Lucy Pickering, Texas A & M University, Commerce	
Spoken French in a pronunciation course: Impressions and applications	217
Jessica Miller, University of Wisconsin, Eau Claire	
Frédérique Grim, Colorado State University	
Integration of pronunciation in first-year German textbooks	229
Iulia Pittman, Auburn University	
Implications of English as a Lingua Franca for pronunciation teaching in teacher education	243
Veronika Thir, University of Vienna	
Assessing assessment: A principled revision of an in-house pronunciation diagnostic test	253
Patricia Watts, University of Illinois at Urbana-Champaign	
Amanda Huensch, University of South Florida	

Teaching Tips

Walk __ By: Raising learner consciousness about unstressed words	262
Donna Brinton, Educational consultant	
Pronunciation Workout!	267
Marsha Chan, Sunburst Media and Mission College	
Video voiceovers for helpful, enjoyable, pronunciation practice.....	270
Lynn Henriksen, Brigham Young University	
Feel the rhythm! Fun and effective pronunciation practice using Audacity and sitcom scenes	277
Edna Lima, Ohio University	
A communicative approach and dialect exposure enhance pitch accent awareness by learners of Japanese	285
Vance Schaeffer, Indiana University	
Isabelle Darcy, Indiana University	

A STEP FURTHER

John Levis (jlevis@iastate.edu), Rania Mohammed (rhani@iastate.edu) Manman Qian (mqian@iastate.edu), Ziwei Zhou (ziweish@iastate.edu)
Iowa State University

The 6th Annual Pronunciation in Second Language Learning and Teaching Conference took place in Santa Barbara, California on the beautiful campus of the University of California Santa Barbara. Over 100 participants from 17 countries attended, and most attended both days of the conference, despite the allure of Pacific Ocean beaches only a short walk away. We thank Dorothy Chun, the conference organizer, for putting together a wonderful conference experience. The conference has developed into a meeting place for researchers and teachers from around the world who are looking at second language pronunciation in relation to many different second languages. Top names in the field, new and upcoming researchers, and classroom teachers all address a topic that is close to their hearts. This has been a goal of the conference from the beginning, and it is increasingly evident in the conference line-up. This year included papers, to name a few, about Karen and Vietnamese speakers learning Swedish, Polish learners of English, Multiple L1s learning Norwegian, Russian and Chinese learners of Italian, English learners of Spanish, French and Chinese, Cantonese learners of French, Americans learning Russian, pronunciation in German textbooks, acquisition of Arabic, French learners studying abroad, Greek, Japanese and Spanish vowels in English.

The plenary speaker, Professor Alene Moyer of the University of Maryland, spoke on “Learner Autonomy in Second Language Phonology: Choice vs. Limits.” Her talk is not included in the proceedings because a leading journal asked her to write it up for them. This has happened several times to our plenary speakers, and it reflects the high quality of the plenary presenters and the growing interest in second language pronunciation. (It’s not bad for conference publicity, either.)



Abstract

Learning a new sound system poses challenges of a social, psychological, and cognitive nature, but the learner’s decisions are key to ultimate attainment. This presentation focuses on two essential concepts: *choice*, or how one wants to sound in the target language; and *limits*, or various challenges to one’s goals vis-à-vis accent. Qualitative and quantitative data underscore the relevance of learner autonomy as a guiding principle from which to explore related constructs such as self-determination, motivation, decision-making and self-concept. We also review several prominent *limits* on phonological attainment to counterbalance and contextualize the aspect of *choice*. Suggestions are given for both teaching and research that prioritize learner autonomy with reference to a complexity perspective.

Over twenty-five people helped review papers submitted for the proceedings. Thank you to Amanda Baker, Richard Cauldwell, Isabelle Darcy, Tracey Derwing, Amanda Huensch, Okim Kang, Ettien Koffi, Shannon McCrocklin, Alison McGregor, Jessica Miller, Rania Mohammed, Murray Munro, Lucy Pickering, Mandy Qian, Marnie Reed, Arkadiusz Rojczyk, Viviane Ruellot, Anita Saalfeld, Veronica Sardegna, Sinem Sonsaat, Jessica Sturm, Ron Thomson, Pavel Trofimovich, Elisabeth Zetterholm, Beth Zielinski and Ziwei Zhou.

A final step further for this set of the proceedings is that we now have an official ISSN identifier (ISSN 2380-9566) for the PSLT Proceedings. One of the contributors encouraged us to do this, and she was right. It was past time to do it.

Sessions

Besides the plenary, the conference typically has three equally important types of sessions. First are the 20-minute concurrent sessions. This was the first year that part of the conference went to three concurrent sessions, an indication of the greater number of presentations as the conference becomes better known and attracts researchers from a wide variety of contexts.

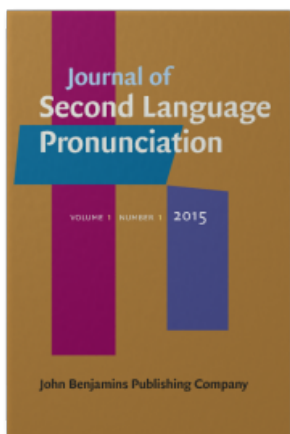
The second type of session is the poster presentations. Posters are given a dedicated time slot and lunch is provided for all participants to encourage excellent attendance. (It works.) The poster sessions are lively, interesting, and could easily take more time than is scheduled. Like oral presentations, poster presentations are often superb in quality. Posters often provide lively interaction with participants as well. In a two-day conference, there are only so many slots for presentations, and poster sessions allow us to provide a wider range of presentations than would otherwise be possible.

The third type of session is the newest, the Teaching Tips Roundtable. On Saturday morning, the second day of the conference, we start with a 90-minute session. Each presenter is given a table to set up, and 8-10 participants join them for a 10-minute demonstration and discussion of an unusual or innovative way to teach pronunciation for some language. After 10 minutes, a bell rings and participants go to another table. During a 90-minute session, presenters meet with different groups 7-8 times, participants get to try out 7-8 new ways to teach pronunciation, and we manage to make concrete connections between theory and practice. This year there were 10 presenters. The Teaching Tips idea comes from a session used by the Speech Communication national conference called "My Great Idea". This was our second year and the popularity of the session for all participants means that it will become a mainstay of the conference. One well-known researcher said many of the technology-oriented teaching tips were far more innovative than her university (known for its use of technology in language learning) had demonstrated.

Other Notable Features

The conference provided a conference dinner, refreshments between sessions, a lunch for the first day, and an opening reception in the tropical gardens of the conference hotel. These pieces are now standard parts of the conference experience and are included with the registration fee.

The conference will also continue to offer freely available, peer-reviewed conference proceedings with their own ISSN number. Because of our desire to provide excellent papers that meet a research-publishing niche just below journals and book chapters but well above the content of newsletters, these proceedings papers are reviewed, revised and proofed carefully before being published. They typically come out before the next conference.



The proceedings are now joined by a new publication that is a direct result of this conference. The new *Journal of Second Language Pronunciation*, published by John Benjamins, began in 2015 and has now had two issues published. The journal provides an identity to this growing field and a scholarly visibility. The relationship between the Proceedings and the Journal was one that we were uncertain about. We were not sure if the proceedings would have served their purpose, but this year's proceedings are the most robust yet, with nearly 30 papers available. It seems clear that there is plenty of room for pronunciation-related research of all kinds, in conference proceedings, newsletters, book chapters, pedagogically-oriented books, and scholarly journals. We are happy to provide these proceedings as a contribution to an ever-growing field.

The conference line-up

The PSLLT Proceedings serve two purposes. They are a serious publishing venue that reflects the work being done in a wide variety of contexts, providing data, findings, and ideas for studies to other researchers, and they are a record of the conference itself, a kind of history for those who did not attend and receive a conference program (or perhaps lost the program they did receive). As a result, we include the final conference line-up next. (The full program, including abstracts of each session, can be found on the [psllt.org](http://psllt.org/index.php/psllt/index/manager/files/PSLLT2014programupdated9-3.pdf) website at <http://psllt.org/index.php/psllt/index/manager/files/PSLLT2014programupdated9-3.pdf>)

Thursday, September 4th

7:00 Reception at Pacifica Suites for Conference Attendees

Friday, September 5th

8:00-8:50	Registration (Humanities and Social Sciences Building)		
9:00-9:10	Welcome (McCune Conference Room)		
9:10-10:10	Plenary Address by Alene Moyer (McCune Conference Room, 6020 HSSB (Humanities and Social Sciences Building))		
10:10-10:30	Break		
	ED 1215 - Perception	ED 1217 – Effects of Instruction	ED 1207 – Teaching
10:30-10:55	The Effect of Listening Context On Native Speakers' Perception of Mandarin Tones (Jiang & Chun)	Contrastive Stress can be learned – But can it be taught at lower levels? (Muller Levis, Levis, & Benner)	Form-focused Pronunciation Activities: To Repeat or Not to Repeat? (Foote & McDonough)
11:00-11:25	The effects of perceptual training on pronunciation (Inceoglu)	Content Effects in Native English-Speaking Students' Adaptation to the Speech of ITAs (Hayes-Harb & Cotsonas)	Music in the pronunciation classroom: Are all approaches created equal? (Barrett)
11:30-11:55	Phonological memory, speech perception and oral fluency (Thomson)	Effecting Change in Pronunciation Practice Behaviors: The Learners' Journey (Sardegna & McGregor)	What to Target in Second Language German Pronunciation Instruction: Findings from the Classroom (Roccamo)

12:00- 1:50	Box Lunch (Provided in Education 1207)
	Posters: ED 1215 and ED 1217

12:30-1:50	<p style="text-align: center;">Poster Titles and Presenters</p> <p>A test of the speech learning model: non-native acquisition of Arabic sounds - Al-Mahmoud</p> <p>Nonnative or Native: Do students in an upper-intermediate EAP course have a preference? - AlShahrani/Chan</p> <p>The acquisition of English L2 prosody by Italian native speakers - Busa/Stella</p> <p>Putting it all together: From pronunciation analysis to pronunciation pedagogy - Crabtree</p> <p>Putting Italian vowels in the mouths of Russian and Chinese speakers - De Meo/Vitale</p> <p>The effects of L1 in the syllabification of French- de Moras</p> <p>The role of pitch contours in teaching vowel length distinctions in Japanese - Deguchi</p> <p>The influence of linguistic stereotyping on grammaticality judgments - Ghanem</p> <p>TV Arabic Speech markers analysis - Heider/Belakova</p> <p>Linguistic factors in the acquisition of connected speech in second language Spanish - Holt</p> <p>The impact of computer-aided pronunciation training on suprasegmental perception and production skills in an ESP program - Jolley/Tanner</p> <p>Pronunciation errors faced by ESL Pahari speakers - Khan</p> <p>Pronunciation Problems of the Pahari EFL Learners: A Case Study of the BS 4 Year Program, University of AJK, Pakistan - Khan/Qadir</p> <p>Learner preferences and the learning of Japanese rhythm - Kinoshita</p> <p>The acquisition of Korean prosody by native English speakers and its role on L2 discourse - Lee</p> <p>When French becomes tonal: Prosodic transfer of L1 Cantonese speakers - Lee/Matthews</p> <p>Acoustic production and perception analysis of Montenegrin English L2 vowels - Lucic</p> <p>Pronunciation features affecting comprehensibility - Margolis</p>
------------	---

	<p>general American accents - Nyemer/Tanner</p> <p>Assessing double phonemic boundary among Brazilian Portuguese learners of English - Osbourne</p> <p>Implementation of Pronunciation in Turkish Classes - Ozcelik Evaluating innovative pronunciation training methods in the Austrian EFL classroom - Pfandl-Buchegger/Insam</p> <p>Integration of pronunciation in first-year German textbooks - Pittman</p> <p>Perception of French Accents of Immersion Graduates in British Columbia - Poljak/Munro</p> <p>Feeding the Beast: First Wave Innovation in Pronunciation Teacher Education - Reed</p> <p>Attitudes toward the teaching of L2 pronunciation among high school Spanish teachers - Reeder</p> <p>Pronunciation Issues for Korean Speakers in Linguistic Perspective - Sands</p> <p>Acquisition of word final devoicing by American learners of English - Simonchyk</p> <p>Teaching of pronunciation in EFL teacher training programs in Turkey - Sonsaat</p> <p>Implications of English as a lingua franca for pronunciation teaching in English language teacher education - Thir</p> <p>Assessing assessment: A pronunciation diagnostic case study - Watts/Huensch</p> <p>IPA training to improve comprehensibility of EFL learners - Yokomoto</p> <p>Swedish Tonal Word Accents - Zetterholm/Tronnier</p> <p>The realization of narrow focus in L2 Chinese - Zhang</p> <p>Examining the effectiveness of teaching English intonation to Brazilian EFL</p>		
	ED 1215 - Technology	ED 1217 - Acquisition	ED 1207 - Teaching
2:00-2:25	Acquisition of L2 Mandarin Chinese Tones with Learner--- Created Tone Visualizations (Chun, Jiang, Meyr, & Yang)	Short-term study abroad: French learners' fluency, comprehensibility and accentedness (French, Gagné, Guay, & Beaulieu)	What's Hot? What's Not? Insights from Pronunciation Practitioners (Brinton & Chan)

2:30-	Expanding CALST: multilingual analysis	Long-term effects of early foreign language	Where we've been and where we're
2:55	of L1-L2 phonotactics for language teaching (Martinez-Paricio, Koreman, Husby, Abrahamsen & Bech)	learning on phonemic discrimination (Harada)	going: 1850-2014 and beyond (Murphy & Baker)
3:00- 3:25	The Potential of ASR for Fostering Pronunciation Learners' Autonomy (McCrocklin)	The Effect of Visual Feedback on VOT Productions by L2 Learners of Spanish (Offerman & Olson)	Relationship between L2 Oral Communication Strategies and Abilities (Hardison)
3:30- 3:55	Break		
	ED 1215 – L1/L2 issues	ED 1217 – Effects of Instruction	ED 1207 - Descriptive
4:00- 4:25	The memory of the mother tongue on the second language articulation of affricates (Insung Ko)	Spoken French in a Phonetics Course: Impressions and Applications (Grim & Miller)	Revisiting the Pronunciation of English by Speakers from Mainland China (Huang & Pickering)
4:30- 4:55	Automatizing the principled identification and extraction of minimal pairs (Qian)	L2 Graduate Students' Development of Speaking in Oral Presentations (Buss & Kennedy)	Mexican & Chicano Spanish intonation: Differences related to Information Structure (Miglio)
5:00- 5:25	The Girlfriend or the Girl's Friend: Karen Speakers' L1 attribute transfer to L2s—a conversational problem. (Zetterholm)	The effects of peer tutoring on speech production and perception outcomes for learners of Spanish (Pak)	The production of word boundary C#V sequences in English by Polish learners (Rojczyk, Schwartz & Balas)
6:00	Conference Dinner		

Saturday, September 6th

8:30-9:00	Registration (ED 1207)	
9:00-10:30	Teaching Tips Round Robin (ED 1215 and 1217)	
	<p>Teaching French Language and Culture Through Phonetics (Colantonio)</p> <p>A communicative approach and dialect exposure enhance pitch accent awareness by learners of Japanese (Schaefer, Darcy)</p> <p>Video voiceovers for helpful enjoyable pronunciation practice (Henrichsen)</p> <p>Introducing a suite of high-interest communicative activities for providing learners' concentrated exposure to English Word Stress (Richards)</p> <p>Walk ___ By: Raising learner consciousness about unstressed words (Brinton)</p> <p>Pronunciation workout! (Chan)</p> <p>Using portable mobile technologies for shadowing activities (Foote) Feel the rhythm! Fun and effective pronunciation practice using <i>Audacity</i> and sitcom scenes (Lima)</p> <p>Automatic Speech Recognition: Making it Work for your Pronunciation Class (McCrocklin)</p> <p>The Sandwich Approach: The Secret to Attaining Unconscious Competence (Reed)</p>	
10:30-10:55	Break	
	ED 1215 – Teachers & Students	ED 1217 - Intelligibility
11:00-11:25	Windows of Opportunity for L2 Fluency and Pronunciation Development (Munro & Derwing)	Online pronunciation instruction: Improving speaker comprehensibility and enhancing learning experience (Lima)
11:30-11:55	Can explicit instruction promote the development of French L2 learners' socio-phonetic knowledge? (Beaulieu & French)	Pronunciation Features of Intelligible Speech Among Different Varieties of World Englishes (Kang, Moran,

		Thomson)
12:00-12:30	Do Oral Corrective Feedback Practices Differ Between ESL and EFL Teachers? (Tanner & Bashford)	Impressionistic and Instrumental Account of the Intelligibility of [θ] in L2 English (Koffi)
12:30-2:00	Lunch (not provided)	
	ED 1215 – New approaches	ED 1217 - Teaching
2:00-2:25	Expanding the vowel space: A corpus investigation of native speakers of Greek, Japanese, & Spanish (Sakai)	A NAIL in the Coffin of Stress-Timed Rhythm (Dickerson)
2:30-2:55	The Role of Intonation in the Organization of L2 Academic Discourse (Buss, Cardoso, & Kennedy)	Adult student's perspectives on the benefits of pronunciation instruction (Henrichsen)
3:00-3:25	Statistical Learning in L2 Phonology: Production of Word-Final Stop Codas in L2 English (Vokic)	The evaluation of the pronunciation component in a large scale ESP curriculum (Sheppard)
3:30-3:55	A comprehensive analysis of word stress of new general service list vocabulary (Richards)	Liaison in L2 French: the effects of instruction (Sturm)
4:00-4:25	Closing (ED 1217)	

The Proceedings

With so many papers, there are multiple ways to organize them, and some papers fit nicely into more than one category. We have divided them into sections: Experimental Studies (8), Technology and L2 Pronunciation (4), Instructional Issues and Approaches (11), and Teaching Tips (5). In each category we have organized them alphabetically. A summary of each paper is included below.

Experimental Studies

In “Acquisition of English L2 prosody by Italian native speakers: Experimental data and pedagogical implications,” Maria Grazia Busà and Antonio Stella investigate Yes-No question intonation patterns in English L2, Italian L1, and English L1. Their study shows that grammatical functions may be communicated by different intonation patterns, and intonation choices are subject to both transfer of L1 categories and the acquisition of L2 phonological categories.

Larissa Buss, Walcir Cardoso, and Sara Kennedy explore the ways in which intonation is used to organize speech (“Discourse intonation in L2 academic presentations: A pilot study”). They look at how four L2 graduate students employed intonation as an organizational device during their first six months in the L2 environment, comparing them to native English-speaking controls. The two intonational features they looked at were *paratones* (extra high pitch at the beginning of a new discourse topic) and *mid-utterance pitch boundaries* (which should indicate the connection between phrases with non-low pitch). Overall, L2 participants marked mid-utterance pitch boundaries adequately, but paratones were considerably less noticeable than the native speakers.

In “Mexican & Chicano Spanish prosody: Differences related to information structure,”

Michael J. Harris, Viola G. Miglio, and Stefan Th. Gries look at how monolingual and bilingual heritage Spanish speakers use intonation in naturalistic language to express given and new information. They expected that bilingual speakers (because of their experience with English) would use pitch more to encode information structure, a hypothesis confirmed by their analysis.

Naoko Kinoshita looks at the effects of different teaching techniques for introducing rhythm to learners of Japanese (“Learner preference and the learning of Japanese rhythm”). Multiple instruction techniques were introduced during Japanese pronunciation classes, which were 1) rhythm marking, 2) clapping, 3) grouping rhythmic patterns, 4) haiku, 5) pronouncing to a beat, 6) visual acoustic analysis (Praat), and 7) shadowing. Students showed strong preferences for particular techniques. In addition, the learners improved their ability to hear minimal pairs differing in rhythmic structures.

In a study of the interaction of French and Cantonese, Jackso Lee and Stephen Matthews (“When French becomes tonal: Prosodic transfer from L1 Cantonese and L2 English”) look at the effect of an L1 tone language (Cantonese) and L2 non-tone language (English) on the acquisition of an L3 non-tone language (French). It appears that syllables of French content words (nouns, verbs,

adjectives, and so forth) all are prosodically similar to the Cantonese high level tone, whereas syllables of French function words (prepositions, determiners, etc.) carry a different Cantonese tone (low level). This study recognizes the potential complications of transfer from different languages to a new language.

In “Are French immersion “accents” unique?” Livia Poljak looks at whether or not a French immersion program resulted in accents that were distinct from native French speakers. Results were significant, but it was not always easy for native French listeners to distinguish the two accents. Longer utterances produce more accurate choices.

Ala Simonchyk and Isabelle Darcy, in “Acquisition of word-final devoicing by American learners of Russian,” investigated word-final devoicing by American learners of Russian. Learners did not fully demonstrate word-final devoicing in Russian even though voiceless consonants are unmarked and occur in English word-finally. This suggests that transfer is not given even when the feature exists in both L1 and L2.

Finally, In “Swedish tonal word accents produced by Vietnamese L1 speakers,” Elisabeth Zetterholm and Mechthild Tronnier look at how speakers of Vietnamese (a tone language) distinguish tonal patterns in L2 Swedish. Although some studies suggest that speakers of tonal languages may have an advantage in learning other tonal languages, others have seen no advantage. Results revealed that one of the two Swedish tonal patterns (which is similar to a Vietnamese tone) was more successfully used than the other (which was not similar).

Technology and L2 Pronunciation

In “Audiovisual and auditory-only perceptual training: Effects on the pronunciation of French nasal vowels,” Solène Inceoglu investigated whether perceptual training on the three French nasal vowels led to improvement in vowel production. The study also looked at whether improvement differed based on treatment. Results showed that both training groups significantly improved, but that the audiovisual training group improved significantly more than the production of the audio-only training group.

The Norwegian research group of Violeta Martínez-Paricio, Jacques Koreman, Olaf Husby, Jardar Eggesbø Abrahamsen, and Øyvind Bech discuss computer-assisted pronunciation training for speakers of multiple L1s learning L2 Norwegian (“Consonant clusters in online L2 teaching: A multilingual approach”). They report on the development of a multilingual database that allows comparison of the phoneme inventories of a large number of languages and a computer-assisted pronunciation training system which uses contrastive analysis to select different pronunciation exercises depending on the native language of the user.

In “Extracting minimal pairs automatically with word frequency and phonetic environment controlled: Introducing a program written in PERL,” Manman Qian reports on the development a

computational tool that automates minimal-pair selection for pronunciation practice, controlling for L1 background, word frequency and syllable environment. Minimal pairs can be generated for English learners from 22 different native language groups.

Shannon McCrocklin writes about the usefulness of commercial speech recognition for pronunciation training in “Automatic speech recognition: Making it work for your pronunciation class.” This paper introduces the benefits of using ASR, explores differences in some readily available technologies, gives ideas for using ASR in a pronunciation class and discusses challenges to using ASR programs.

Instructional Issues and Approaches

In “A sociocultural view of engagement in the music-based pronunciation classroom,”

Catrice Barrett situates her study in the Sociocultural Theory framework that sees language learning as a process of mediated cognition where students’ behaviors are regulated from person to person. Through analyzing students’ engagement in a collaborative rap activity that focused on word stress, she shows that students are able to appropriate linguistic resources to regulate their own learning trajectory to meet specific language objectives, while attending to and resolving discrepancies in language abilities.

“What’s hot, what’s not: Insights from pronunciation practitioners” (Donna Brinton and Marsha Chan) reports on key discussions on pronunciation issues from an online discussion list for pronunciation researchers and practitioners. They summarize several extensive discussions of different topics from the list over the course of a year.

Janay Crabtree, in “Putting it all together: From pronunciation analysis to pronunciation pedagogy,” is concerned with how research on intelligibility, comprehensibility, and accentedness is internalized by TESOL teacher trainees (TTs) and translated into pronunciation classrooms. By having TTs analyze speech samples and write lesson plans, the study finds ambiguous responses from TTs towards features of intelligible, comprehensible, and accented speech. The author concludes by calling for more rigorous and vigorous training of TTs.

In “The role of pitch contours in teaching vowel length distinctions in Japanese,” Masanori Deguchi explores more reliable cues to distinguish vowel length for Japanese learners. Since the traditional “quantitative” distinction of vowel length has posed difficulties for learners of Japanese, the author argues for a focus on a “qualitative” distinction – pitch contour – to serve as a more reliable cue. A new way for teaching students how to distinguish vowel length is also suggested.

Attending to the aged old debate of classifying English as stress-timed language, Wayne Dickerson (“A nail in the coffin of stress-timed rhythm”) proposes an alternative two-peak

profile model to account for the English rhythm patterns. The author argues that rather than standing on three pillars, the two-peak profile is attested by empirical evidence and should be considered as the foundation of learners' spontaneous speech.

Lynn Henrichsen and Christa Stephens, in "Advanced adult ESL students' perspectives on the benefits of pronunciation instruction," investigate students' perceptions and perspectives on the effects of pronunciation instruction. The report confirmed previous studies that students' comprehensibility and accented did not exhibit significant change. However, the qualitative analysis points to the perceived benefits of pronunciation instruction. Specifically, students felt that their awareness, confidence, listening skills, and strategy use generally improved.

In "Revisiting the pronunciation of English by speakers from Mainland China," Meichan Huang and Lucy Pickering look at the English pronunciation difficulties of Chinese learners from an underrepresented area in China. The paper argues that Chinese learners face varied difficulties based on region, and that Chinese learners should not be treated as an undifferentiated group.

Jessica Miller and Frédérique Grim ("Spoken French in a pronunciation course: Impressions and applications") examined how students enrolled in university French pronunciation courses feel about standard and non-standard varieties of French in the second language classroom. Most felt that standard and non-standard varieties were important in the classroom, both for perception and production.

In "Integration of pronunciation in first-year German textbooks," Iulia Pittman looks at how pronunciation is integrated into popular first-year German textbooks used in American colleges. A review of the top ten first-year German textbooks found that only half of the textbooks include any information on pronunciation and only 20% of the textbooks present it in an effective way, a finding that suggests that German pronunciation pedagogy has room for significant improvement.

Veronika Thir revisits issues with ELF pronunciation in teacher training programs, in "Implications of English as a Lingua Franca for pronunciation teaching in teacher education." ELF pronunciation remains a topic of perennial interest but there is little evidence that it has significantly influenced pronunciation teacher training. This paper suggests ways to help future NNS teachers gain the knowledge and skills necessary to make informed decisions for English pronunciation teaching in a globalized world. The paper argues that pronunciation training, theoretical education, and critical reflection are all essential in helping NNS teachers to develop a positive professional identity as English pronunciation teachers.

An approach to practical pronunciation assessment is presented in "Assessing assessment: A principled revision of an in-house pronunciation diagnostic test" (Patricia Watts and Amanda Huensch). The authors report on the process involved to evaluate and revise an existing

pronunciation diagnostic test. Modifications included an added section testing aural perception, changing the free speech section from a self-introduction to an interview, and revising the targeted segmentals based on principled selection criteria.

Teaching Tips

In “Walk __ By: Raising learner consciousness about unstressed words,” Donna Brinton provides a way to help learners hear reduced unstressed elements in English speech.

“Pronunciation Workout!” (Marsha Chan) is just what it sounds like: a physical workout focused on pronunciation and articulation. The teaching tip has many links to videos to provide something of the experience of the live teaching tip.

In “Video voiceovers for helpful, enjoyable, pronunciation practice,” Lynn Henrichsen shows how to use videos from animated film scenes to extend the benefits of rehearsed pronunciation practice and tracking. The teaching tip provides links to videos demonstrating the technique in practice.

Edna Lima (“Feel the rhythm! Fun and effective pronunciation practice using Audacity and sitcom scenes”) demonstrates ways to practice perception and production of rhythm in English using *Audacity* and sitcom scenes. The combination of the two technologies allows the use of authentic-like spoken models and the development of carefully-constructed practice materials.

Finally, in “A communicative approach and dialect exposure enhance pitch accent awareness by learners of Japanese,” Vance Schaeffer and Isabelle Darcy demonstrate a number of traditional and unusual techniques for the teaching of Japanese pronunciation to English speakers.

Busà, M. G., & Stella, A. (2015). The acquisition of English L2 prosody by Italian native speakers: Experimental data and pedagogical implications. In J. Levis, R. Mohammed, M. Qian & Z. Zhou (Eds). *Proceedings of the 6th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Santa Barbara, CA (pp. 15-26). Ames, IA: Iowa State University.

THE ACQUISITION OF ENGLISH L2 PROSODY BY ITALIAN NATIVE SPEAKERS: EXPERIMENTAL DATA AND PEDAGOGICAL IMPLICATIONS

Maria Grazia Busà, University of Padova, Italy

Antonio Stella, University of Salento, Italy

This paper investigates Yes-No question intonation patterns in English L2, Italian L1, and English L1. The aim is to test the hypothesis that L2 learners may show different acquisition strategies for different dimensions of intonation, and particularly the phonological and phonetic components. The study analyses the nuclear intonation contours of 4 target English words and 4 comparable Italian words consisting of sonorant segments, stressed on the semi-final or final syllable, and occurring in Yes-No questions in sentence-final position (e.g., *Will you attend the **memorial**?*, *Hai sentito la **Melania**?*). The words were contained in mini-dialogues of question-answer pairs, and read 5 times by 4 Italian speakers (Padova area, North-East Italy) and 3 English female speakers (London area, UK). The results show that: 1) different intonation patterns may be used to realize the same grammatical function; 2) different developmental processes are at work, including transfer of L1 categories and the acquisition of L2 phonological categories. These results suggest that the phonetic dimension of L2 intonation may be more difficult to learn than the phonological one.

INTRODUCTION

Prosody, that is, the variations in rhythm, stress, pitch, and intonation patterns occurring during speech, has an important function in language and communication. It is used to signal emphasis, disambiguate sentences, contextualize meaning, as well as convey emotions and attitudes.

L2 speakers' prosody is likely to be heavily influenced by the L1, with effects on L2 production and perception. In fact, L2 prosodic production has been shown to contribute to what is perceived as 'foreign accent', and to have an impact on L2 speakers' comprehensibility and intelligibility. For example, the perception of L2 fluency and speech seems to be affected by differences between L1 and L2 stress, speech rate and timing, pitch and intonation (Anderson-Hsieh et al., 1992; Kormos & Dénes, 2004; Munro & Derwing, 2001; Munro, 2008; Wennerstrom, 2000).

L2 prosodic production also affects the pragmatics of communication. For example, L2 prosodic production may result in disfluency, overlaps or interruptions in the flow of speech; it may thus contribute to making conversation unpleasant or frustrating, or cause distraction or annoyance (e.g., Anderson-Hsieh et al., 1992; Munro, 2008). Also, L2 prosodic production may be associated with paralinguistic meaning that is not intended by L2 speakers, but that may form the basis for prejudice or negative stereotyping, and thus increase the odds of L2 speakers' social or professional discrimination (see studies reviewed in Busà, 2007). For example, Northern

Standard Germans' (NSG) speech tends to be associated with lack of liveliness by Southern Standard British English (SSBE) speakers because it is characterized by a rather low pitch. In comparison, SSBE speakers sound overexcited to NSG speakers because SSBE is characterized by very varied pitch (Mennen et. al., 2008).

Finally, L2 prosodic production may result in speakers' inability to mark information structure in discourse, that is, to assign prominence to speech constituents. This may make L2 speakers unable to differentiate between lexically stressed and unstressed elements in a sentence, signal emphasis, prioritize information, disambiguate sentences etc. (Mennen, 2007; Ramírez Verdugo, 2006).

The importance of prosody for successful communication in L2 is unquestioned, and its role is generally recognized in today's L2 pronunciation teaching courses (e.g., Romero-Trillo, 2012; Trouvain & Gut, 2007). However, prosody is notoriously considered difficult to teach and learn, due to its inherent complexity and the intricate relations between linguistic and paralinguistic features that characterize it (Wrembel, 2007). Teaching prosody is also made difficult by the fact that L2 learners are not always aware of the uses and meanings of prosody even in their own language. To overcome this, it has been suggested that teaching prosody should involve raising learners' awareness about prosody and intonation patterns (Chun, 2002; Wrembel, 2007)

Despite the importance of prosody to L2 communication, we still know little about L2 prosody acquisition processes and the effects of L2 prosodic production on listeners. More comparative descriptions of L1 and L2 prosodic systems are needed to gain knowledge about L2 prosodic production, and to provide data for enhancing teaching methods and materials. Such accurate descriptions are often missing, or present single and standardized accounts of prosodic phenomena –though, in fact, prosody is characterized by considerable variation that is traceable to differences between speakers, language variety, linguistic function, paralinguistic meaning, etc. (e.g., Clopper & Smiljanic, 2011; D'Imperio, 2002; Grabe, 2004; Joerg et al., 2002; Kelly 2012).

This study investigates the nuclear intonation contours of English Yes-No questions produced by Italian native speakers, and compares them with those produced in English L1 and Italian L1. In so doing, it aims to contribute to the field of L2 acquisition by providing data on English L2 prosodic production.

About L2 prosody acquisition

One of the active strategies in L2 prosody acquisition is transfer, a well-known mechanism also in the acquisition of L2 segments. In fact, it has been shown that prosodic transfer may occur even after learners have been exposed to the L2 for considerable time (Trofimovich & Baker, 2007; Mennen, 2007).

For segments, current models (SLM (Flege, 1995) and PAM/PAM-L2 (Best, 1995; Best & Tyler, 2007)) have proposed that L2 acquisition processes and outcomes may be affected by the degree of similarity/dissimilarity between L1 and L2 sounds, which would affect learners' perception by making them identify some of the L2 sounds with some of their native language sounds (Strange, 2007).

It is likely that a principle of similarity/dissimilarity may apply also in the acquisition of prosody. However, for prosody, this principle would seem to be much harder to establish than for

segments, due to the complex nature of prosodic systems, the variability existing across speakers and dialects, and the interrelations between forms and linguistic and paralinguistic functions (e.g. Ladd, 1996; Gussenhoven, 2006).

In order to generate predictions as to the relative difficulty of production and perception of L2 intonation based on a scale of cross-language similarity/dissimilarity, Mennen (2015) has proposed that similarities and differences between L1 and L2 intonation should be characterized along four dimensions (Mennen, 2015: §3.1). These are: (1) The phonological (or systemic) dimension, which concerns the inventory of prosodic elements that are linguistically meaningful in the language (e.g., types of intonation patterns, pitch accents, prosodic words, etc.), as well as the typology of combinatorial possibilities (of tune-text associations, tone sequences, etc.); (2) The phonetic (or realizational) dimension, which concerns how the prosodic features are implemented –for example, the scaling (i.e., the relative height) of pitch accents in an utterance, their alignment with the segments, or their shape (i.e., steep falling or rising); (3) The semantic dimension, which concerns how the prosodic elements are used for conveying meaning –for example, crosslinguistic differences in the use of pitch for marking focus or interrogativity (e.g., while English marks yes/no questions with rising intonation, Greek uses falling intonation (Mennen, 2015)); (4) The ‘frequency’ dimension, which relates to similarities and differences in the cross-linguistic use of phonological elements, e.g., pitch accents and boundary tones.

According to Mennen (2015) L2 learners may show deviations from the native norm in each of the four dimensions of intonation, but the deviations occurring in some dimension may be more frequent than others, probably because not all intonation dimensions present the same level of difficulty for L2 learners. Some evidence also suggests that different dimensions may be subjected to different acquisitional rules, in the sense that over time learners may improve in some dimensions of intonation but not others. For example, studies show that L2 learners may eventually be able to acquire the phonology, but may rarely acquire the phonetics of L2 prosody (Mennen, 2007; Stella, 2012; Ueyama, 1997).

However, it remains to be determined which aspects of L2 prosody are more difficult for learners to acquire.

The Present Study

This paper investigates nuclear intonation contours in Yes-No questions in English L2, as produced by Italian native speakers, and in Yes-No questions in Italian L1 and English L1 as baseline NS data. The aim is to test the hypothesis that L2 learners may show different acquisition strategies for different dimensions of intonation. In particular, this study investigates the acquisition of the phonological and phonetic components of intonation. Considerations will be drawn on the effects of L2 learners’ intonation for perception/communication and pedagogy.

METHODS

Materials, Subjects and Procedure

The experiment compares the production of English L2 speech by Italian native speakers with the production of Italian L1 and English L1 speech as baseline NS data.

One test (English L2) and two baseline NS (English L1 and Italian L1) data sets were created for

the experiment; from each data set Yes-No questions were extracted from mini-dialogues of question-answer pairs read by the subjects.

The questions contained 4 highly comparable target words having all-sonorant segments, and stress on the antepenultimate or the penultimate syllable. These words are, for English: **Memorial**, **Banana**, **Normandy**, **Memory**; and for Italian: **Melania**, **Banane**, **Lamina**, **Mobile** (stressed syllables are indicated in bold). The words occurred in sentence-final position in each data set (e.g., *Will you attend the **memorial**?*, *Hai sentito la **Melania**?*).

The subjects were 4 Italian speakers (Padova area, North-East Italy) and 3 English female speakers (London area, UK). All speakers had a similar level of competence in English (B2, Common European Framework of Reference for Languages (CEFR)). The Italian speakers read the English L2 and the Italian L1 data sets; the English speakers read the English L1 data set. Each speaker read the data set 5 times at a normal pace. Thus, the material consists of 80 tokens for English L2 and Italian L1 (4 target words x 5 repetitions per 4 speakers), and 60 tokens for English L1 (4 target words x 5 repetitions per 3 speakers). The speakers were recorded in a quiet room with the professional equipment available at the *Language and Communication Lab*, University of Padova.

Phonetic and phonological analysis of the data

The phonological analysis focused on the nuclear pitch accent of the target words. The acquired audio signals were labeled using the *Praat* software (Boersma & Weenink, 2009). We labeled the onset and offset of all syllables in the accented word and the sequence of high (H) and low (L) tones of the pitch accent and the boundary tone associated with that word. Tonal targets are labeled HN and LN for nuclear pitch accents and HE and NE for boundary tones. The pattern of alternation of H and L tones was then used to classify the whole F0 contour and to determine the occurrence of each intonational pattern out of the total amount of the speakers' productions.

The phonological description of pitch accents occurring in nuclear position was conducted auditorily and through a phonetic examination of tones alignment, i.e. measuring the ratio between the latency of the H and L tonal targets and the duration of the associated tonic syllable. We used this measure as a criterion to determine the phonological category to assign to accents. An Univariate ANOVAs and Tukey post-hoc tests were then run on data to test whether our phonological description is stable among speakers. Alternatively, a t-tests for independent samples is run when referring to two speakers only.

This data analysis aimed at providing a phonological description of the nuclear contours that occur more frequently in Yes-No questions produced in English L2 and Italian L1 by native (Padova) Italian speakers, and Yes-No questions produced in English L1 by (London) English native speakers.

RESULTS

Italian L1 Nuclear Intonation Contours

In Italian Yes-No questions, the Padova Italian speakers produce nuclear intonation contours with falling pitch accents (70%), as well as rising pitch accents (30%). Both accents are followed by either rising (48.7%) or falling boundary tones (51.3%). Figure 1 shows the phonetic

realization of falling accent and rising accent in two productions ending with a rising boundary tone. For each contour, the ToBI phonological label assigned to the accent is shown.

The rising accent is labeled L+H* (Fig. 1, top panel) and represents the typical realization of the Padua Italian variety –the use of the rising accent in nuclear position of Yes-No questions is highly marked with respect to other varieties of Italian (Gili Fivela et al. 2015). This accent is aligned with the tonic syllable in the production of the 4 Italian speakers: LN is aligned at a mean of 18% of the tonic syllable with a significant difference among the 4 speakers ($F(3,20)=6.814$; $p=0.002$); HN is aligned at 94% without any difference between the speakers ($F(3,20)=2.256$; $p=0.113$).

The falling accent, labeled H+L* (Fig. 1, bottom panel), is another option available to the Padova Italian speakers and is a more common pattern for Yes-No questions among other varieties of Italian (Gili Fivela et al., 2015). The results show that the F0 fall is synchronized with the tonic syllable: the means show that H is aligned 6% before the onset of the tonic syllable, and L is aligned at the end of the syllable. Both targets are significantly different among speakers (H: $F(3,52)=11.153$; $p=0.000$; L: $F(3,52)=22.562$; $p=0.000$). However the post-hoc tests show that 3 out of 4 speakers show the same alignment: for H only speaker 2 produces a different pattern, while for L it is speaker 4 who produces a different pattern.

English L1 Nuclear Intonation Contours

The English L1 speakers produced Yes-No questions mainly with two types of nuclear pitch accents: rising accent (63.3%), and high accent (30%). Both are followed by low-level boundary tones.

The rising accent in nuclear position (Fig. 2, top panel) was mostly produced by 2 of the 3 speakers. The F0 rise is consistently synchronized with the tonic syllable: L is aligned at 12% of the tonic syllable, with a significant difference between the two speakers ($t=3.588$; $p=0.001$), and H is aligned at 83%, with no statistical difference between speakers ($t=1.912$; $p=0.064$). On the contrary, the third speaker produces a high pitch accent in nuclear position, with no F0 rise and realized through a high tone sustained from the beginning of the intonational phrase until the half of the tonic syllable (44%). In this case the high tonal target is very difficult to detect because F0 falls gradually along the tonic syllable and the entire contour shows a very narrow pitch range. Given these realizations, the rising accent is labeled L+H* and the high accent H*. While the H* accent is not found in Italian, the L+H* accents are present also in Padova Italian but have a different phonetic form, with L starting from a higher F0 value.

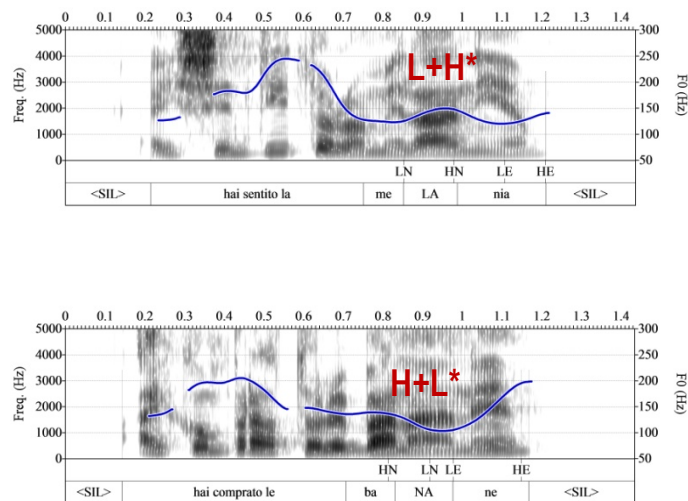


Figure 1. Italian L1 Yes-No question intonation contours. Top panel: Rising accent L+H*. Bottom panel: Falling accent H+L*.

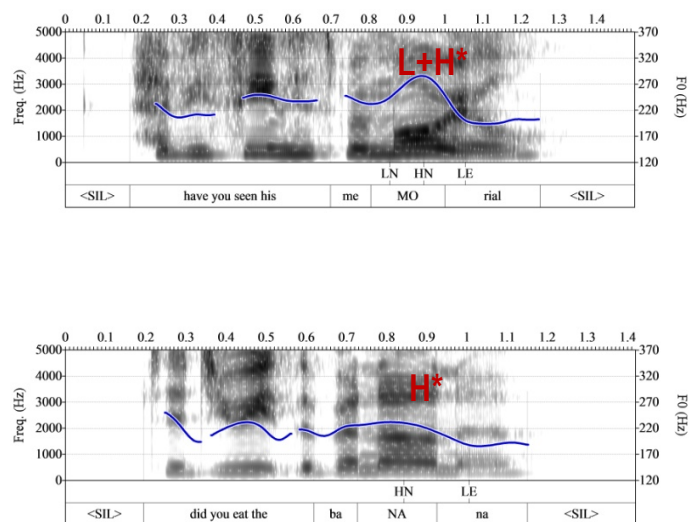


Figure 2. English L1 Yes-No question intonation contour. Top panel: Rising accent L+H*. Bottom panel: High accent H*.

English L2 Nuclear Intonation Contours

The English L2 Yes-No questions show a different distribution and typology of nuclear intonation contours with respect to those found in Italian L1. Examples are provided in Figure 3.

The most frequent pattern is a falling pitch accent H+L* (45%; Fig. 3, top panel), followed by a rising boundary tone; this contour is produced mainly by 2 of the 4 speakers (1 and 4) and represents a clear example of phonological transfer of intonation from L1 to L2: the nuclear accent has indeed a similar auditory result of the H+L* contour produced in Italian L1, as well as the same alignment with respect to the tonic syllable (H: -15%; L: 75%) with more variability between the two speakers (H: $t=9.066$; $p=0.000$; L: $t=7.196$; $p=0.000$).

The H* accent is the second most frequent pattern (27.5%; Fig. 3, central panel) and is produced by the other two speakers (2 and 3). This nuclear accent does not occur in Italian L1, and its realization is highly comparable to the patterns found in English L1. While the phonological forms of English L1 and L2 pitch accents are similar, their implementations look different. In particular, the English L2 accents show high pitch excursion values, and are similar to pitch accents used in Italian L1 but are different from those produced by native English speakers. This can be interpreted as a sign that the Italian speakers have correctly acquired the L2 phonological category to produce the English nuclear accent. However, the transfer of phonological features from the L1 to L2 may affect the phonetic implementation of pitch accents, which might be harder to produce by L2 speakers.

The same is also true for the third kind of nuclear pitch accent, which is produced by speakers 2 and 3 with a similar frequency as the H* accent (26.2%; Fig. 3, bottom panel). This accent is auditorily very similar to the H* pitch accent produced in English L2, but shows an alignment comparable to the L+H* rising accent of English L1 (L: 30%; H: 91%). However, its phonological status is not clear, since the entire nuclear contour shows the same high pitch span and range that characterize the productions in English L2 by native Italian speakers. So, here too, the data show that the phonetic implementation of the pitch contour is heavily influenced by the native language system.

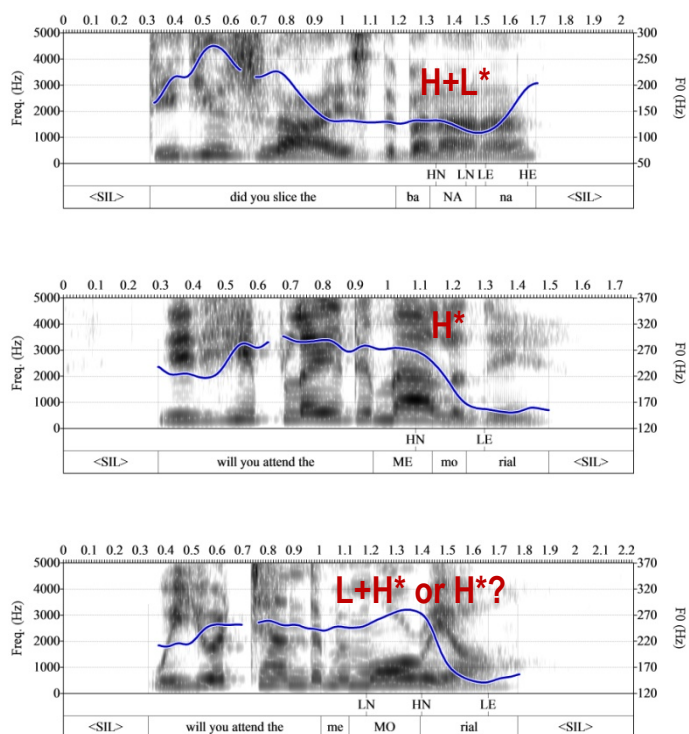


Figure 3. English L2 Yes-No question intonation contours. Top panel: Falling accent H+L*. Central panel: High accent H*. Bottom panel: Rising accent L+H* or H*.

DISCUSSION

This study aimed to investigate the differences in the production of nuclear intonation contours in Yes-No questions in English L2 by Italian native speakers, and in English L1, and Italian L1 for comparison. In doing so, the study also aimed to widen the knowledge on the complex and largely unexplored field of L2 prosody acquisition, a necessary step for the development of methods and materials for teaching L2 prosody to learners.

The first finding is that both L1 and L2 speakers appear to use different types of nuclear intonation structures to express the same grammatical function, confirming previous findings that intonation is characterized by considerable intra- and inter-speaker variation (D'Imperio, 2002; Gili Fivela et al., 2015; Grabe, 2004; Ladd, 1996). It should be observed that for English L1, the present data does not coincide with previous descriptions of intonation patterns observed in London English L1 (Grabe, 2004). This could be further indication of the variability existing in (London) English, or could be due to differences in experimental procedures. This would need further investigation.

The second finding is that the development of English L2 intonation by Italian learners seems to be characterized by the co-existence of different strategies and types of acquisition processes. The most frequent strategy is the transfer of intonation patterns from the L1, evidenced by the

existence of nuclear contours that are similar, with regard to both phonological type and phonetic details, to those of the L1. However, the learners do not transfer in the L2 the L1 pattern that they identify as most typical of their L1 (i.e., L+H*), showing that they are using some phonological filter in the transfer of categories from L1 to L2.

Secondly, the learners can produce at least some aspects of L2 intonation accurately: this is evidenced by the existence in the L2 of a nuclear contour that is present in English but not in Italian, i.e., H*. However, this contour is produced with pitch ranges that are typical of the L1 and not of the L2, which means that, by failing to implement the phonetic details of the L2 intonation patterns, learners lag behind in the phonetic/realizational dimension.

The present data provide support to Mennen's (2015) proposal that not all dimensions of L2 intonation are acquired equally, and that some dimensions may present L2 learners with greater difficulty than others. It would appear that the phonological dimension is easier to learn for the Italian learners of English than the phonetic dimension, confirming previous research carried out on other L1s and L2s (Mennen, 2007; Stella, 2012; Ueyama, 1997), and showing that L2 learners may be able to acquire the phonology, but rarely the phonetics of L2 prosody.

This study was not designed to test the effect of differences between L1 and L2 in the acquisition of L2 intonation along the semantic and frequency dimensions. However, it is highly possible that such differences play an important role in the development of L2 intonation by contributing to determining what learners perceive as perceptually similar/dissimilar (Mennen, 2015). Future investigations will test these effects, and try to understand the role of all four dimensions of intonation (Mennen, 2015: §3.1) in determining the observed variability.

Objects of further investigation will also be the effect of experience, as a function of age of learning, period of acquisition, and frequency of use of L2 vs. L1. All these factors are known to have a large effect on the L2 acquisition process, and would also contribute to explaining the variability observed in the present data.

Ultimately, as suggested by Mennen (2015), studies of L2 intonation should be aimed at predicting, among other things, the relative difficulty of L2 intonation categories, causing divergences in L2 learners' intonation patterns: what makes these categories difficult to learn, how do they affect learners' acquisition development, along which dimensions do they appear to be more evident, etc.

Being able to predict and recognize what causes divergences in L2 intonation patterns will help us teachers make generalizations, as well as develop methods and materials that are useful for the students to learn L2 prosody.

ABOUT THE AUTHORS

M. Grazia Busà (mariagrazia.busa@unipd.it) is associate professor of English language and linguistics at the University of Padova, Italy, where she teaches in the Communication Studies Degree program. Her research interests include foreign accent, language acquisition, experimental phonetics and phonology, intercultural communication. Her current research focuses on L2 speech intelligibility, non-verbal communication in L2, and the pragmatic outcomes of accented speech and non-verbal language in multilingual and multicultural communication.

Antonio Stella (antonio.stella@unisalento.it) obtained a Ph.D. in Linguistics (Curriculum in Experimental Phonetics) at the University of Salento (Lecce, Italy) in December 2012 with a dissertation on the articulatory analysis of intonation in non-native German productions. From October 2011 to September 2013 he worked as Research Fellow in English Phonetics at the University of Padova (Italy) on a project on the acoustic analysis of English L2 intonation by (Veneto) Italian speakers. His main research interests are intonational phonology, laboratory phonology, and the analysis of the articulatory correlates of speech using the electromagnetic articulography.

REFERENCES

- Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning*, 42, 529–555.
- Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.) *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171–232). Timonium, Maryland: York Press.
- Best, C. T., & Tyler, M. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.) *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13–34). Amsterdam: John Benjamins Publishing.
- Boersma, P., & Weenink, D. (2009). Praat: doing phonetics by computer [Computer program]. <http://www.praat.org/>.
- Busà, M. G. (2007). New perspectives in teaching pronunciation. In A. Baldry et al. (Eds.), *From DIDACTAS to eCoLingua: An ongoing research project on translation and corpus linguistics* (pp. 171–188). Trieste: EUT.
- Chun, D. (2002). *Discourse intonation in L2. From theory and research to practice*. Amsterdam: John Benjamins.
- Clopper, C. G., & Smiljanic, R. (2011). Effects of gender and regional dialect on prosodic patterns in American English. *Journal of Phonetics*, 39, 237–245.
- D’Imperio, M. (2002). Italian intonation: An overview and some questions. *Probus*, 14, 37–69.
- 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, Maryland: York Press.
- Gili Fivela, B., Avesani, C., Barone, M., Bocci, G., Crocco, C., D’Imperio, M., Giodano, R., Savino, M., & Sorianello, P. (to appear). Varieties of Italian and their intonational phonology. In S. Frota & P. Prieto (Eds.), *Intonational variation in Romance*. Oxford: OUP.

- Grabe, E. (2004). Intonational variation in urban dialects of English spoken in the British Isles. In P. Gilles & J. Peters (Eds.) *Regional variation in intonation* (pp. 9-31). Tuebingen: Niemeyer, Linguistische Arbeiten.
- Gussenhoven, C. (2006). Experimental approaches to establishing discreteness of intonational contrasts. In S. Sudhoff et al. (eds.) *Methods in empirical prosody research* (pp. 321-334). Berlin/New York: De Gruyter.
- Kormos, J., Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, 32, 145-164.
- Joerg, P., Gilles, P., Auer, P. Selting, M. (2002). Identification of regional varieties by intonational cues. An experimental study on Hamburg and Berlin German. *Language and Speech*, 45(2), 115-139.
- Kelly, N. (2012). *Production and Perception of Intonation in Southern American English and West Coast American English*. University of Texas at Austin, Experiments in Linguistics, Spring issue.
- Ladd, R. D. (1996). *Intonational phonology*. Cambridge: Cambridge University Press.
- Mennen, I. (2007). Phonological and phonetic influences in non-native intonation. In J. Trouvain & U. Gut (Eds.). *Non-native prosody. Phonetic description and teaching practice* (pp. 53-76). The Hague: Mouton de Gruyter.
- Mennen, I. (2015). Beyond segments: towards a L2 intonation learning theory (LILt). In E. Delais-Roussarie, M. Avanzi & S. Herment (Eds.) *Prosody and languages in contact: L2 acquisition, attrition, languages in multilingual situations* (pp. 171-188). Berlin: Springer Verlag.
- Mennen, I., Schaeffler, F., & Docherty, G. (2008). A methodological study into the linguistic dimensions of pitch range differences between German and English. In P. A. Barbosa, S. Madureira & Reis, C. (eds.) *Proceedings Of The Speech Prosody 2008 Conference* (pp. 527-530). May 6-9, 2008, Campinas, Brazil.
- Munro, M. J. (2008). Foreign accent and speech intelligibility. In J. G. Hansen Edwards & M. L. Zampini (Eds.). *Phonology and second language acquisition* (pp. 193-218). Amsterdam: John Benjamins.
- Munro, M. J., & Derwing, T. (2001). Modelling perceptions of the comprehensibility and accentedness of L2 speech: The role of speaking rate. *Studies in Second Language Acquisition*, 23, 451-468.
- Ramirez Verdugo, M. D. (2006). Prosodic realization of focus in the discourse of Spanish learners and English native speakers. *Estudios Ingleses de la Universidad Complutense*, 14, 9-32.
- Romero-Trillo, J., Ed. (2012). *Pragmatics, prosody and English language teaching*. Dordrecht: Springer.
- Stella, A. (2012) *Allineamento articolatorio dei toni nella produzione dell'intonazione L2*. Unpublished Ph.D. Thesis. Università del Salento, Lecce (Italy).

- Strange, W. (2007). Cross-language phonetic similarity of vowels: theoretical and methodological issues. In O.-S. Bohn, and M. J. Munro (Eds.). *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 35-55). Amsterdam: John Benjamins.
- Trofimovich, P., & Baker, W. (2007). Learning prosody and fluency characteristics of L2 speech: The effect of experience on child learners' acquisition of five suprasegmentals. *Applied Psycholinguistics*, 28, 251- 276.
- Trouvain, J., & Gut, U., Eds. (2007). *Non-native prosody. Phonetic description and teaching practice*. The Hague: Mouton de Gruyter.
- Ueyama, M. (1997). The phonology and phonetics of second language intonation: The case of "Japanese English." In *Proceedings of the 5th European Speech Conference*, (pp. 2411-2414). Rhodes, Greece.
- Wennerstrom, A. (2000). The role of intonation in second language fluency. In H. Riggensbach (Ed.) *Perspectives on fluency* (pp. 102-127). Ann Arbor, MI: University of Michigan Press.
- Wrembel M. (2007). Metacompetence-based approach to the teaching of L2 prosody: Practical implications. In J. Trouvain & U. Gut (Eds.). *Non-native prosody. Phonetic description and teaching practice* (pp. 189-209). The Hague: Mouton de Gruyter.

Buss, L., Cardoso, W., & Kennedy, S. (2015). Discourse intonation in L2 academic presentations: A pilot study. In J. Levis, R. Mohammed, M. Qian & Z. Zhou (Eds). *Proceedings of the 6th Pronunciation in Second Language Learning and Teaching Conference* (ISSN 2380-9566), Santa Barbara, CA (pp. 27-37). Ames, IA: Iowa State University.

DISCOURSE INTONATION IN L2 ACADEMIC PRESENTATIONS: A PILOT STUDY

Larissa Buss, Concordia University

Walcir Cardoso, Concordia University

Sara Kennedy, Concordia University

Intonation is an important cue to the organization of oral discourse, being used by first language (L1) English speakers to signal topic shifts and relationships among parts of speech (Wennerstrom, 1994, 1998). This pilot study investigated how four second language (L2) graduate students' use of intonation as an organizational device developed naturalistically during their first six months immersed in an L2 environment. The participants were recorded delivering four short presentations at approximately two-month intervals. The recordings were analyzed for two features: *paratones* (extra high pitch at the beginning of a new discourse topic) and *mid-utterance pitch boundaries* (which should indicate the connection between phrases with non-low pitch). Two presentations given by native English speakers were also analyzed for comparison. Overall, the L2 participants adequately produced non-low pitch boundaries within utterances, but their F0 peaks at topic shifts were considerably less prominent than those observed in the native-speaker data. Three participants' use of intonation changed over time, either improving on the features analyzed or displaying a U-shaped pattern of development.

INTRODUCTION

Intonation can have a variety of communicative functions, which include conveying the illocutionary force of an utterance, expressing attitudes and emotions, and building rapport (Chun, 2002; Pickering, 2001). No less important is its role in organizing discourse (Wennerstrom, 1998). In English, variations in pitch are used alongside syntactic and lexical devices to create a hierarchical structure in spoken discourse, highlighting important information and signaling relationships among constituents. However, given that intonation patterns vary across languages, it may take time, exposure, and practice for learners to master the subtleties of discourse intonation in their second language (L2). This pilot study analyzed changes in L2 graduate students' use of intonation as an organizational device, specifically paratones and utterance-medial pitch boundaries, over two academic terms.

Description of the Intonation Variables

Intonation is an important cue to the organization of large stretches of speech. A new discourse topic usually begins with a high peak on the first prominent syllable and a high pitch overall, followed by "a series of lesser peaks" (Yule, 1980, p. 36). When the topic comes to an end, there is a lowering of the pitch level and often a slowing down, followed by a lengthy pause (Yule, 1980). These cues segment discourse into macro-units that are analogous to paragraphs in

writing. Brown (1977) referred to them as *paratones*, but they have also been called *phonological paragraphs*, *pitch sequences*, *sequence chains*, and *intonational paragraphs*.

Besides marking the macro-structure of discourse, intonation is also used to indicate the relationship between parts of speech. A low pitch boundary, observed at the end of declarative statements, is generally associated with finality or closure (Wennerstrom, 2001). It indicates that the speaker has finished what they wanted to say. In an interaction, a marked fall in pitch (low) signals the speaker's intention to give away the turn. On the other hand, high-rising, low-rising, partially falling, and plateau boundaries (all referred to in this study as "non-low") indicate interdependency between the phrases in which they occur and those that follow. That is, they anticipate subsequent phrases for their full and correct interpretation (Wennerstrom, 2001).

Research on the Intonation Variables

Researchers have reported similar findings on how native speakers use paratones and pitch boundaries and on the phonetic/phonological cues associated with these features. Swerts and Geluykens (1994) analyzed the features in the context of spontaneous monologues in Dutch. It was found that the fundamental frequency (F_0) peaks of topic-introducing clauses were higher than the F_0 peaks of the following clauses. Moreover, there was a significant association of low-ending contours with topic finality and of high-ending contours with non-finality. Thompson (2003) described the use of paratones to signal the larger-scale organization of authentic and pedagogically prepared academic lectures in English. The paratones were characterized by extra high pitch on the first prominent syllable and extra low pitch on the last, sometimes ending with decreased volume or speed, laryngealization, and a long pause.

Grosz and Hirschberg (1992) found evidence that listeners pick up on these intonational cues. Seven listeners were asked to label the discourse structure of a news story, either from text alone (with most punctuation removed) or from text and speech. For both modalities, phrases labeled as initiating discourse segments were produced with a larger pitch range than other utterance-initial phrases. In another study (Lehiste, 1975), a group of 30 listeners were asked to determine the context of production of randomized recorded sentences (if isolated, paragraph-initial, medial or final). The results revealed that sentences produced with high F_0 peaks were consistently identified as being paragraph initial. This was interpreted as evidence that listeners have certain expectations regarding the intonation of sentences within an orthographic paragraph, which is a macro-unit comparable to the paratone.

In L2 speech, listener expectations related to discourse intonation are not always met. Tyler, Jefferies, and Davies (1988) observed that the intonation contours of Korean and Chinese teaching assistants (TAs) sometimes ended in a falling pitch, even within sentences, which obscured markings of continuation and finality. According to the authors, this also made the speakers sound abrupt, cold, and impatient. Likewise, Wennerstrom (1994) found that Japanese and Thai subjects used low pitch boundaries in the middle of utterances, failing to establish the link between semantically related phrases. Furthermore, they did not increase pitch range to signal new topics like L1 English and Spanish speakers did. Similar distinctions between native and nonnative speakers were reported in Pickering (2004). In this study, American TAs were found to consistently use the phonological cues of paratones to organize their discourse, whereas the same cues in the speech of Chinese TAs often did not correspond to semantic or structural boundaries in their discourse.

There is some evidence that inaccurate use of discourse intonation may undermine effective communication. Tyler and Davies (1990), for example, described an interaction between a Korean TA and an undergraduate student where communication failed due to problematic intonational cues. Furthermore, Wennerstrom (1998) found a significant relationship between paratone accuracy and the speaking scores of L1 Mandarin speakers lecturing in L2 English. The author also noticed that speakers with high scores tended to use non-low intonation in utterance-medial boundaries, as did native speakers, whereas the participant who received the lowest score produced many more misleading low boundary tones.

Despite the importance of paratones and pitch boundaries, little is known about how these features are acquired by L2 learners immersed in an English-speaking environment. Previous research suggests that learners can improve their pronunciation even without explicit instruction, especially in their first year of immersion (Derwing, Thomson, & Munro, 2006; Flege, 1988). However, it is not clear whether this applies to features of discourse intonation, and, if so, how these features develop over time in L2 speech. The present study contributes to our understanding of how discourse intonation develops naturalistically in L2 English speakers during their first six months immersed in an L2 context. The genre chosen for analysis was the academic presentation.

Research Questions

There are four research questions in this study, two related to paratones and two related to pitch boundaries:

1. Do L2 English/L1 Mandarin speakers differ from L1 English speakers in their use of paratones in oral presentations? If so, how?
2. Do L2 English/L1 Mandarin speakers differ from L1 English speakers in their use of utterance-medial pitch boundaries in oral presentations? If so, how?
3. Does the L2 speakers' use of paratones change across two semesters of immersion in an academic, English-speaking environment? If so, how?
4. Does the L2 speakers' use of utterance-medial pitch boundaries change across two semesters of immersion in an academic, English-speaking environment? If so, how?

METHODS

Participants

This study is part of a larger project examining longitudinal changes in the speaking skills of nonnative graduate students at a Canadian university. The participants here were four L1 speakers of Mandarin who were living abroad for the first time and who had arrived in Canada approximately one month before data collection. They did not receive any formal instruction on pronunciation or academic speaking during the 26-week study. More information is given in Table 1, with the participants' names changed to preserve confidentiality.

Table 1

Participant Information

Name	Gender	Age	Test Score	Program
Xu	F	22	IELTS 7.0	Graduate Diploma in Instructional Technology
Yang	F	25	TOEFL 92	Ph.D. in Civil Engineering
Wei	M	23	IELTS 6.5	M.A.Sc. in Mechanical Engineering
Chen	M	28	TOEFL 83	Master of Applied Computer Science

Two Canadian native English speaker graduate students were also included in the study as controls, a male aged 30 (here called Peter) and a female aged 39 (here called Kim). They were unaware of the purpose of the study.

Data Collection

Speech samples from the L2 graduate students were collected four times over a period of 26 weeks at approximately two-month intervals. A few days before each data collection session, the participants were asked to prepare a five-minute presentation explaining a key concept or term in their field to an imagined audience of first-year undergraduate students. The two native controls were recorded giving in-class presentations for a course. Each of them gave one presentation of approximately five minutes where they summarized and commented on part of a research paper that they had read.

Coding and Analysis

The presentations were transcribed and coded for intonational phrases, phonological utterances, and discourse topics. The first two constructs were conceptualized following Nespor and Vogel's (1986) model of prosodic phonology. According to this model, the intonational phrase (*I*) is the domain of an intonation contour, and its boundaries coincide with the places where pauses may be introduced in an utterance. An *I* is delimited by the boundaries of a root sentence (Emonds, 1976), which is roughly equivalent to a main clause or independent clause—one which is not embedded within a larger structure. Nonetheless, a root sentence does not necessarily form a single *I*. The division of a string into *I*s varies as a function of several factors, including length, rate of speech, emphasis, and style.

Besides a coherent intonation contour and pauses, other phonetic/phonological cues that helped to identify the boundaries of intonational phrases were: lengthening of the final syllable, acceleration at the beginning of a new phrase, a change in the pitch level or pitch direction of an unaccented syllable, and the presence of at least one pitch accent in the intonation phrases created (Cruttenden, 1986).

The phonological utterance (*U*) is the largest constituent in Nespor and Vogel's (1986) hierarchy, and it is formed by one or more intonational phrases. It usually corresponds to the string dominated by the highest node of a syntactic tree, or what is generally recognized as a "sentence". Although *U* can be defined using syntactic information, phonological and semantic factors may cause the prosodic constituent to be restructured, such that the end result may not necessarily be isomorphic to any syntactic constituent.

The pitch boundaries of *Is* situated in the middle of a *U* (i.e., in non-final position) were the aspects of interest in this study. These phrases were expected to end in non-low pitch boundaries, which express "non-finality" and dependency on subsequent phrases (Wennerstrom, 1998). Following Wennerstrom (1998), the pitch boundaries were classified as "low" or "non-low" based on auditory impression and observation of F_0 contours on Praat (Boersma & Weenink, 2014). The final F_0 value of the contour was also noted to determine whether the pitch dropped to the bottom of the speaker's pitch range (Wennerstrom, 2001). Finally, the percentage of non-low boundaries was calculated for each presentation. Truncated (i.e., uncompleted) intonational phrases and fragments that did not contain at least one pitch accent, such as fillers occurring in isolation (e.g., 'uh', 'eh', 'and', 'so'), were excluded from the analysis.

Discourse topics were identified mainly on the basis of semantic content, as "an aggregate of coherently related events, states, and referents" (Chafe, 1994, p. 121). Coding was also aided by the presence of "macro-markers" (Chaudron & Richards, 1986) such as "The first thing I'm going to look at is..." and "The second thing we need to do is...". When used, these markers made the discourse structure very explicit. Yet, most of the time, the speakers transitioned directly from one topic to another or signaled the shift using "micro" discourse markers such as "Also", "As well", and "Then" (Chaudron & Richards, 1986).

Based on the method used in Swerts and Geluykens (1994), the analysis of the paratone was done by measuring the first F_0 peak in each paratone-initial and paratone-final *I* on Praat (i.e., the first F_0 peak in the first and last *Is* of each paratone). Then, the value found for the final *I* of one paratone was subtracted from the value found for the initial *I* of the subsequent paratone. This resulted in measures of pitch increase at rhetorical junctures. Each participant's measures were converted into percentages and averaged.

RESULTS

Paratones

The analysis of F_0 peaks at the beginning of paratone-initial and paratone-final *Is* revealed considerably higher percentages of pitch increase for the L1 controls than for L2 speakers. As shown in Figures 2 and 3, the native speakers increased their pitch by 48% (Peter) and 57% (Kim) to signal topic shifts, whereas the nonnative speakers' increases ranged from 7% to 32%, with a total average of 19%. Large differences between native and nonnative speakers were also observed in Wennerstrom (1994, 1998). As in Wennerstrom (1998), the female control produced a higher increase than the male, but this gender difference was not found for the L2 participants. Figure 1 shows an example of pitch increase from the end of topic 7 to the beginning of topic 8 in Peter's presentation.

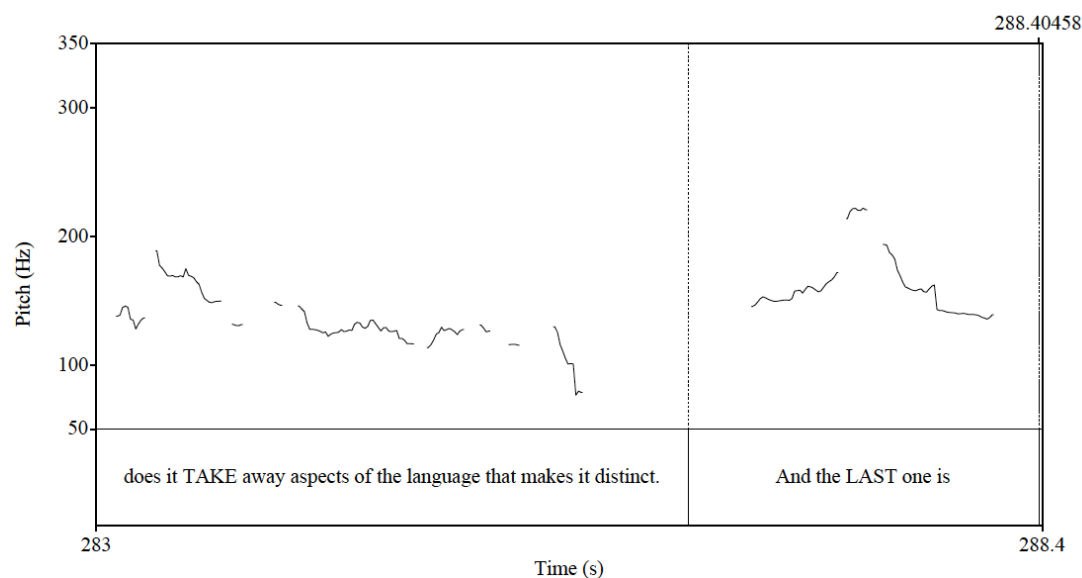


Figure 1. Pitch increase at a topic shift in Peter's presentation (L1 English).

Longitudinally, results seemed to differ for male and female L2 speakers. Xu and Yang, who had higher proficiency test scores than the males (see Table 1), started the study with low percentages of pitch increase at P1, but seemed to improve over time (see Figure 3). As displayed in Figure 4, Wei and Chen started at higher percentages of pitch increase than the females and closer to their L1 counterpart. However, Chen's means dropped at P2 and P3, then increased again at P4. Wei, on the other hand, did not exhibit any major changes.

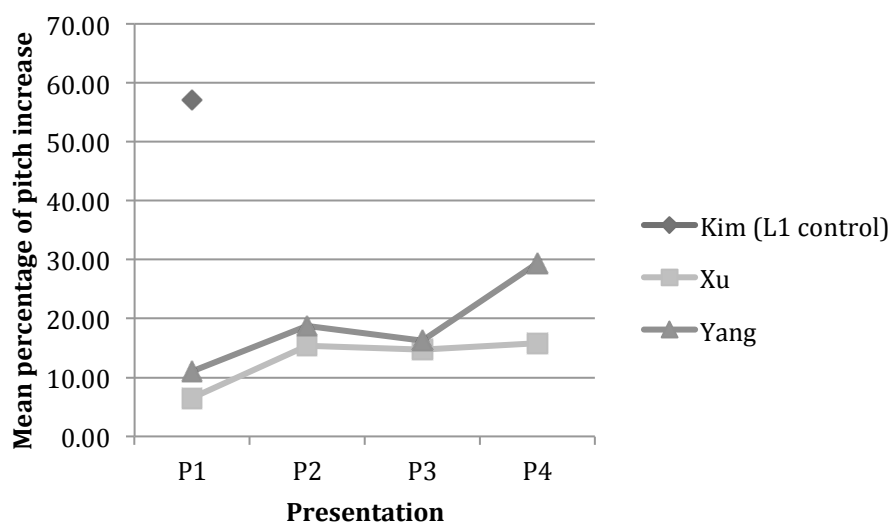


Figure 2. Mean percentage of pitch increase at topic shifts for female participants.

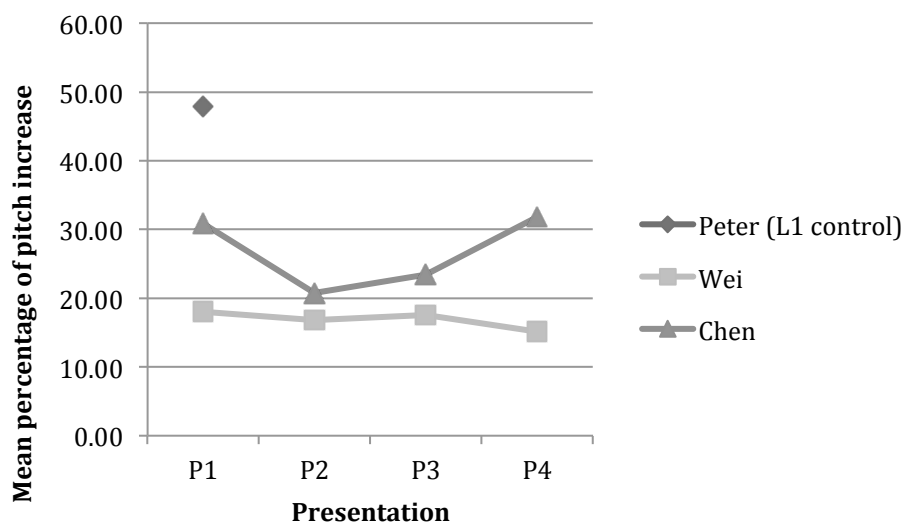


Figure 3. Mean percentage of pitch increase at topic shifts for male participants.

Utterance-Medial Pitch Boundaries

In the native speaker presentations, all mid-utterance boundaries were non-low, as displayed in Table 2. This feature was almost at ceiling level for L2 speakers as well, as most presentations exhibited over 90% of non-low boundaries in the context analyzed. The only exception was the first presentation by Chen, where 82% of the boundaries were non-low. Figure 4 shows the F0 contour for a problematic utterance produced by Chen in his first presentation (P1). The four pitch boundaries are low (↓), plateau (→), low (↓), and partially falling (↘). The two low boundaries misleadingly indicate closure rather than connection between the phrases.

Table 2

Percentages of Non-Low Utterance-Medial Boundaries

Participant	P1	P2	P3	P4
Peter (control)	100	-	-	-
Kim (control)	100	-	-	-
Xu	91	97	95	97
Yang	92	94	100	99
Wei	97	96	100	100
Chen	82	93	100	100

Over time, Chen improved his use of pitch boundaries, as his percentage of non-low boundaries increased to 93% at P2 and 100% at P3. It is possible that Xu and Yang also experienced slight improvement before reaching the ceiling. Overall, however, the accuracy of utterance-medial boundaries was high, therefore no major differences from the native controls or longitudinal changes were observed.

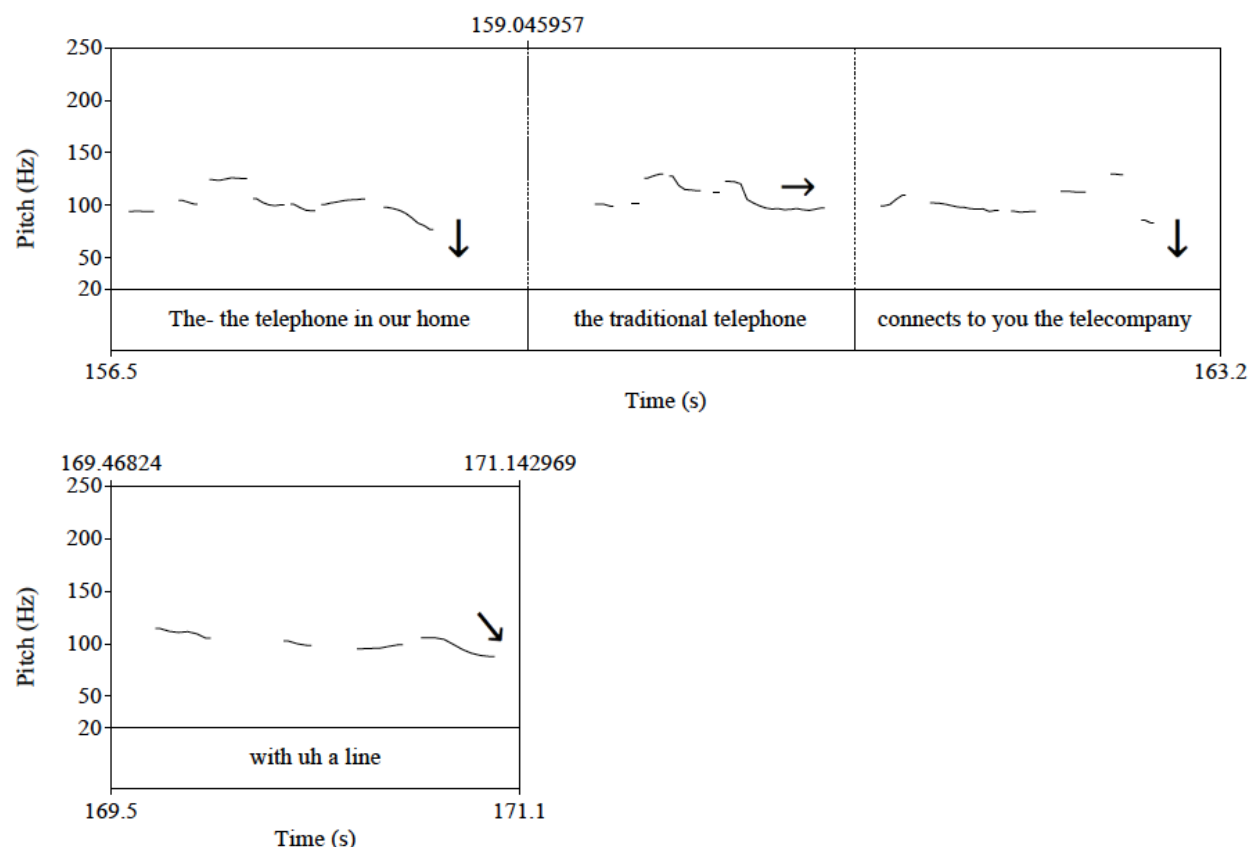


Figure 4. Pitch boundaries in Chen's first presentation.

DISCUSSION

Regarding the first research question, our findings are in line with the evidence that even advanced L2 speakers of English, and particularly L1 speakers of Mandarin, may fall short of native patterns of discourse intonation (Pickering, 2004; Wennerstrom, 1998). All of the L2 participants produced peaks at topic shifts that were considerably lower than those produced by the native controls. Less pronounced peaks at the beginning of new discourse topics might make the macro-structure of a presentation less salient to the audience and could possibly affect comprehensibility, given that previous research has shown a correlation between accurate use of prosody and comprehensibility ratings (Anderson-Hsieh, Johnson, & Koehler, 1992; Wennerstrom, 1998).

In terms of the second research question, the use of pitch at utterance-medial boundaries did not seem to be a challenge for the L2 speakers in this study, as their percentages of non-low boundaries approximated those of the native controls. This suggests that, overall, the participants were able to link *Is* intonationally, correctly signaling the dependency of mid-utterance *Is* on subsequent discourse for their interpretation. This finding differs from Wennerstrom (1998), where 10 out of 16 participants produced less than 80% of non-falling boundaries, possibly because they had lower oral proficiency in English.

The third and fourth research questions asked about potential changes across the two semesters of immersion. Longitudinal changes were observed for three of the four participants, particularly in their use of paratones. However, they did not change in the same way. The two female participants, who had the highest proficiency scores, increased their marking of topic shifts and their use of non-low boundaries over time. Yang showed a larger change in her paratones, as her percentages of pitch increase went up from P1 to P2 and from P3 to P4, whereas Xu plateaued after P2. Changes in utterance-medial pitch boundaries for both Xu and Yang were small due to the ceiling effect. Chen, one of the male participants, became less accurate in his use of paratones from P1 to P3, then improved considerably from P3 to P4, such that his end point was close to his starting point. Interestingly, the opposite happened to his pitch boundaries, which became more accurate over time until ceiling level was reached at P3. No changes were observed in Wei's use of paratones, and his utterance-medial boundaries were almost all non-low from P1, so there was no room for improvement.

Despite not having received any instruction on English pronunciation or academic speaking during the study, Xu and Yang might have improved naturalistically simply by being immersed in the L2 environment or by attending or giving academic presentations. The fact that they did better than the male participants might be due to possibly higher metalinguistic awareness, as previous research has found a positive relationship between language proficiency and awareness (Renou, 2001). Increased awareness may in turn facilitate further language development when the learners arrive in an environment with more opportunities for L2 input and output – and, consequently, more opportunities to notice the gap between the two. In contrast, a learner like Wei, who had lower English proficiency, might require more time in the L2 context, more practice or even explicit instruction in order to make progress.

The U-shaped curve observed for Chen's paratones has been previously reported in the literature for other L2 features, including verbs (Kellerman, 1979) and codas (Abrahamsson, 2003). U-shaped behaviour occurs when structures that are initially correct or closer to target “undergo a process of attrition, only to be reestablished at a later stage” (Sharwood Smith & Kellerman, 1989, p. 220). The learners' start and end point may be similar, but the process of cognitive restructuring that takes place between the two points changes their underlying knowledge of the language. This process might have been triggered when Chen moved to Canada and started to have more exposure to and practice with the language.

Given the pilot nature of this study, findings should be regarded as preliminary only, and no firm conclusions can be drawn at this point. It is still to be verified whether the other participants in the project will show similar development. Admittedly, the speech samples analyzed in this study do not necessarily reflect the speakers' spontaneous use of intonation, as the presentations could be prepared beforehand and were often aided by the use of PowerPoint and written notes. Also, the L2 presentations were not completely authentic in that the audience was student research assistants. This method was chosen, however, to make the presentations more

comparable and to allow participants' presentations to be recorded around the same time. Finally, another limitation was the analysis of only one presentation by each native speaker. Gathering similar longitudinal data from L1 speakers would allow for a better comparison with the L2 presentations and would help to control for task effects.

ABOUT THE AUTHORS

Larissa Buss (larissacks@gmail.com) is a PhD student in Education specializing in Applied Linguistics at Concordia University, in Montreal. Her research interests include teacher education, pronunciation instruction, and second language phonetics and phonology. She has a full scholarship for her doctoral studies from the CAPES Foundation of the Ministry of Education of Brazil.

Walcir Cardoso (Ph.D. McGill University, walcir@education.concordia.ca) is an associate professor of Applied Linguistics (Department of Education, Concordia University – Montreal, Canada). He conducts research on the second language acquisition of phonology, morphosyntax and vocabulary, and the effects of computer technology (particularly clickers, text-to-speech synthesizers, and automatic speech recognition) on L2 learning.

Sara Kennedy (sara.kennedy@education.concordia.ca) teaches at Concordia University in Montreal, Canada. She conducts research on the effects of instruction and exposure on listening and speaking in a second language, and on the production, comprehension, and assessment of L2 speech. Her current research projects involve immigrants and post-secondary students who are learning and studying in a second language.

REFERENCES

- Abrahamsson, N. (2003). Development and recoverability of L2 codas: A longitudinal study of Chinese-Swedish interphonology. *Studies in Second Language Acquisition*, 25(3), 313–349.
- Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning*, 42(4), 529–555.
- Boersma, P., & Weenink, D. (2014). Praat: Doing phonetics by computer [Computer program]. Version 5.3.77, retrieved 18 May 2014 from <http://www.praat.org/>.
- Brown, G. (1977). *Listening to spoken English*. London: Longman.
- Chafe, W. (1994). *Discourse, consciousness, and time*. Chicago: University of Chicago Press.
- Chaudron, C., & Richards, J. C. (1986). The effect of discourse markers on the comprehension of lectures. *Applied Linguistics*, 7(2), 113–127.
- Chun, D. M. (2002). *Discourse intonation in L2: From theory and research to practice*. Amsterdam and Philadelphia: John Benjamins.
- Cruttenden, A. (1986). *Intonation* (2nd ed.). Cambridge: Cambridge University Press.
- Derwing, T. M., Thomson, R. I., & Munro, M. J. (2006). English pronunciation and fluency development in Mandarin and Slavic speakers. *System*, 34(2), 183–193.

- Emonds, J. (1976). *A transformational approach to English syntax: Root, structure-preserving, and local transformations*. London: Academic Press.
- Flege, J. E. (1988). Factors affecting degree of perceived foreign accent in English sentences. *Journal of the Acoustical Society of America*, 84(1), 70–9.
- Grosz, B., & Hirschberg, J. (1992). Some intonational characteristics of discourse structure. *Proceedings of ICSLP 92*, Banff, 429–432.
- Kellerman, E. (1979). The problem with difficulty. *Interlanguage Studies Bulletin*, 4, 27–48.
- Lehiste, I. (1975). The phonetic structure of paragraphs. In A. Cohen & S. G. Nooteboom (Eds.), *Structure and process in speech perception* (pp. 195–203). Berlin: Springer-Verlag.
- Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris Publications.
- Pickering, L. (2001). The role of tone choice in improving ITA communication in the classroom. *TESOL Quarterly*, 35(2), 233–255.
- Pickering, L. (2004). The structure and function of intonational paragraphs in native and nonnative speaker instructional discourse. *English for Specific Purposes*, 23(1), 19–43.
- Renou, J. (2001). An examination of the relationship between metalinguistic awareness and second-language proficiency of adult learners of French. *Language Awareness*, 10(4), 248–267.
- Sharwood Smith, M., & Kellerman, E. (1989). The interpretation of second language output. In H. W. Dechert & M. Raupach (Eds.), *Transfer in language production* (pp. 217–36). Norwood, NJ: Ablex.
- Swerts, M., & Geluykens, R. (1994). Prosody as a marker of information flow in spoken discourse. *Language and Speech*, 37(1), 21–43.
- Thompson, S. E. (2003). Text-structuring metadiscourse, intonation and the signalling of organisation in academic lectures. *Journal of English for Academic Purposes*, 2(1), 5–20.
- Tyler, A., & Davies, C. (1990). Cross-linguistic communication missteps. *Text & Talk*, 10, 385–411.
- Tyler, A., Jefferies, A., & Davies, C. (1988). The effect of discourse structuring devices on listener perceptions of coherence in non-native university teacher's spoken discourse. *World Englishes*, 7(2), 101–110.
- Wennerstrom, A. (1994). Intonational meaning in English discourse: A study of non-native speakers. *Applied Linguistics*, 15(4), 399–420.
- Wennerstrom, A. (1998). Intonation as cohesion in academic discourse. *Studies in Second Language Acquisition*, 20, 1–25.
- Wennerstrom, A. (2001). *The music of everyday speech*. New York: Oxford University Press.
- Yule, G. (1980). Speakers' topics and major paratones. *Lingua*, 52, 33–47.

Harris, M., Miglio, V., & Gries, S. Th. (2015). Mexican and Chicano Spanish prosody: Differences related to information structure. In J. Levis, R. Mohammed, M. Qian & Z. Zhou (Eds). *Proceedings of the 6th Pronunciation in Second Language Learning and Teaching Conference* (ISSN 2380-9566), Santa Barbara, CA (pp. 38-48). Ames, IA: Iowa State University.

MEXICAN & CHICANO SPANISH PROSODY: DIFFERENCES RELATED TO INFORMATION STRUCTURE

Michael J. Harris, University of California, Santa Barbara

Viola G. Miglio, University of California, Santa Barbara

Stefan Th. Gries, University of California, Santa Barbara

This study addresses the intonational encoding of new and given information by monolingual Mexican speakers of Spanish and Spanish/English Chicano heritage speakers. Spanish is a so-called non-plastic language, which tends to encode novel information in a speech signal with word order. Meanwhile, English is known as a plastic language, which uses pitch excursions to signal new information. This study compares the acoustic correlates of information structure in a naturalistic corpus of semi-directed interviews in order to evaluate dialectal variation in the prosodic encoding of new information. It was hypothesized that bilingual speakers would use more pitch excursions for new information due to the fact that they also speak a plastic language, namely English. The results conclude that bilingual Chicano speakers do in fact use more plastic, or English-like, pitch excursions to encode new information, as compared to the monolingual speakers. This study is novel in its use of naturalistic language, rather than experimental tasks in examining information structure and in its use of a mixed-effects model to verify the results.

INTRODUCTION

In order to ease the processing load of parsing a speech signal, speakers resort to different strategies to mark novel information in discourse. In general, languages typically use two ways to signal novel information- either prosodic cues or structural (i.e. word-order) changes to the sentence. This paper addresses the manner in which two dialects of Spanish signal new information. The dialects examined are monolingual Mexican Spanish and bilingual Chicano Spanish. These dialects make for an interesting comparison as the Mexican speakers speak what is known as a *non-plastic language*, i.e. a language that uses word order manipulations, rather than pitch excursions to signal new information. Meanwhile, in addition to Spanish, the Chicano speakers also speak English, which has been described as a *plastic language*, using prosodic cues, intonation in particular, to signal new information. This comparison allows a perspective on the effect of English on the prosodic system of Chicano bilingual Spanish, which could explain general differences in pronunciation compared to monolingual Spanish speakers. The following section will introduce the topic of information structure, followed by a description of the methods used in the current paper, including participants, data, and the statistical evaluation. The results and a discussion of the implications of this study conclude the paper.

Information Structure

In speech, contents can be divided into *new information* and *given information*. New information is content that is novel to the discourse (akin to the linguistic notion of narrow focus), while given information is content that has already been mentioned in the discourse (akin to the linguistic notion of broad focus). In order to facilitate the parsing of speech for the listener, a speaker will often mark new information, making it more salient, while given information is not similarly marked. Two different strategies are typically used to draw the listener's attention to new information- intonation and word order. Those languages that are said to use the former strategy are called *plastic languages*. For instance, Dutch and English use intonation to mark information status (Vallduví, 1992). Languages that use the latter strategy are called *non-plastic languages*. These languages generally have more variable word order, and thus employ syntactic rather than prosodic cues in the signaling of new information. Languages such as Catalan and Spanish are non-plastic languages (Cruttenden, 1993; Zubizarreta & Nava, 2010).

While Spanish has been described as a non-plastic language, there are few studies about intonation of different varieties of Spanish, although the articles collected in Prieto and Roseano (2010) and Butragueño (2004) do analyze several aspects of intonation in several dialects of Spanish. There are especially few studies about prosodic correlates of information structure in Spanish (e.g. Cruttenden, 1993), and none, as far as we know are based on naturalistic speech; past studies on information structure are based on controlled elicited or read sentences. However, the use of spontaneous speech as opposed to experimental settings provides data that are closer to actual language usage; natural recorded speech is more likely to reflect spoken prosodic patterns since "[i]t is well known that there are differences between read and unscripted speech" (Deterding, 2001, p. 220). Furthermore, previous literature about pitch and information structure in Spanish (e.g. Cruttenden, 1993) fails to address dialectal variation or the effects of bilingualism (an aspect also criticized in Arvaniti & Garding, 2007 for English). In comparison, the current study addressed two different dialects of Spanish and the effects of English on the prosodic systems of Chicano bilingual speakers.

Research Questions

As mentioned above, work on prosodic encoding of new and given information has been carried out mostly with controlled elicited or read utterances. However, we extend current research by investigating naturalistic speech. The current study investigates monolingual Mexican Spanish and bilingual Chicano Spanish in their prosodic encoding of information structure. Our research questions were as follows:

- Would we find evidence of dialectal variation between monolingual Mexican speakers and bilingual Chicano speakers in their use of prosody to signal new information using a corpus of semi-directed interviews?
- Given our study of heritage speakers – i.e. early bilinguals – would we find that these speakers use pitch to signal new information, as is customary in English (but supposedly not in Spanish)?
- What prosodic variables and/or interactions are relevant in the encoding of information structure by Spanish speakers in general?

As mentioned, monolingual Mexican speakers speak a non-plastic language, whereas, bilingual Chicano speakers initially acquire the Spanish prosodic system, but later (usually during primary school) they also acquire the prosodic system of American English, a plastic language. The Chicano Spanish portion of the corpus was comprised of speakers who principally spoke Spanish with their family, only acquiring English at a later age (but always before 8 years of age, although the participants felt that they were dominant in English). It is accepted that the prosodic system is acquired at a very young age. For instance, Nazzi, Bertoncini, & Mehler (1998) showed that infants and neonates can distinguish between their native speech rhythms and non-native speech rhythms. Thus, it is reasonable to assume that the Chicano speakers acquired a complete monolingual Spanish prosodic system. However, as mentioned, they later became dominant in English, a plastic language. The current study hypothesizes that the influence of the English prosodic system will be detectable in the Spanish of Chicano speakers. That is, due to their use of a plastic language, the Chicano speakers will resort to more pitch excursions in the signaling of new information, as compared to the monolingual Spanish speakers.

METHODS

This section will describe the methodology used to investigate the prosodic encoding of information structure in Chicano and Mexican Spanish. It will describe the participants and data, and the statistical evaluation, including the variables used in the modeling process.

Participants, Data Annotation, and Analysis

The data was taken from a specialized corpus of naturalistic speech compiled for the comparison of bilingual Chicano English and Spanish to various dialects of monolingual English and Spanish. The corpus consists of a series of semi-directed interviews with participants responding to open-ended questions. Each interview is approximately ten minutes in duration. For the current study, the files of 10 monolingual Mexican Spanish speakers (SpeakerType: monolingual) and 10 bilingual Chicano English/ Spanish speakers (SpeakerType: bilingual) were used in the data tagging process. Each group was demographically equivalent. The Mexican speakers consisted of 5 males & 5 females between the ages of 18 and 25 who were born and raised in the greater Mexico City area. These speakers had never extensively studied a language other than Spanish, nor lived in a foreign country. They were all enrolled in a four-year Mexican university at the time of recording. The Chicano speakers also consisted of 5 males & 5 females between the ages of 18 & 25. These were second-generation Californian Spanish speakers (Heritage Speakers). This means that they were born to Mexican parents in the United States who had emigrated from Mexico during or after adolescence and the participants all indicated that they spoke Spanish with their family. These Chicano participants were all enrolled in a four-year Californian university at the time of the recording.

As mentioned, the corpus is comprised of unscripted speech consisting of subjects' narrative responses to recorders' prompts. This study chose to focus on nouns mentioned two or more times by a speaker. From the relevant corpus files, 420 nouns were manually tagged using

Praat (Boersma & Weenink, 2012) for first and subsequent mention(s) of the noun. The data collected for each mention of the noun included a) whether it was the first time the word occurred in the file or a subsequent repetition; b) whether the noun occurred in utterance final

position, c) whether pitch movement (i.e. a pitch excursion) occurred across the word, d) the duration of the stressed vowel, e) maximum intensity, f) what the word was, and g) who the speaker was. Specifically, this tagging involved two types of variables; namely, fixed effects and random effects. Fixed effects are the so-called independent variables or predictors that may be ‘causes’ for behavior of the dependent variable. Random effects account for speaker- or lexically-specific variability in the data by adjusting intercepts of the predictors according to the participant and the word. The tagging was performed by trained linguists in the Phonetics Lab in the Department of Spanish and Portuguese at the University of California, Santa Barbara and is summarized below:

Dependent Variable:

- PitchMovement: if speaker displayed pitch movement across the word, yes vs. no;

Fixed Effects:

- SpeakerType: monolingual vs. bilingual;
- SpeakerSex: male vs. female;
- Givenness: if information is new vs. given;
- LogDuration: the logarithmic transformation of the length of the stressed vowel in ms;
- PhraseFinal: if the word is in final position in the intonational unit (IU), yes vs. no;
- Intensity: peak intensity of the word in decibels (dB);

Random Effects:

- Informant: makes adjustments to intercepts according to the speaker to account for individual variation;
- Word: makes adjustments to intercepts according to each word to account for variation;

Given that the focus of this paper is on the use of tone excursions to signal new information, the inclusion of LogDuration, Intensity, and PhraseFinal in this analysis may seem superfluous as they do not directly address tone or givenness. However, they are necessary. Consider LogDuration and Intensity first. It is well accepted that pitch, duration, and intensity often work in conjunction in naturalistic language. For instance, these cues have been shown to increase for lexical stress (e.g. Fry 1958). Thus, it is important to consider that intensity and vowel duration may increase when tone excursions occur (whether to signal new information or otherwise). Thus, these variables may shed light on differences between the prosodic systems of monolingual and heritage speakers of Spanish. More importantly, these variables provide a more accurate model in that any variation due to intensity and duration are not misidentified by the model as being related to PitchMovement and Givenness. In a similar manner, PhraseFinal is important because duration is known to increase in phrase final position. As prosodic cues often work in union, it is important to control for phrase final position within the data in case tone excursions are more prevalent at the end of an utterance.

After the data tagging process, a generalized linear mixed-effects model selection process was performed. A *linear model selection process* means that interactions between two variables and then individual variables were deleted from a maximal model containing all two-way interactions and variables if they did not significantly contribute to the prediction of the dependent variable, PitchMovement. All statistical analysis was performed using R (R Development Core Team, 2013).

RESULTS

Two predictors were discarded during the model selection process- SpeakerSex and Intensity, leaving a final model with four significant predictors- the three main effects Givenness, LogDuration, PhraseFinal and the significant interaction Givenness : SpeakerType. The final model's performance is intermediately good ($C=0.762$; $marginal R^2=0.22$; $classification\ accuracy=71\%$).

Table 1

Significant Variables and Interactions in Predicting PitchMovement (yes or no)

Significant predictors	Significant interaction
Givenness	Givenness : SpeakerType
LogDuration	
PhraseFinal	

The following sections will discuss the interaction Givenness : SpeakerType, then the two significant predictors that do not participate in this interaction, namely LogDuration and PhraseFinal.

Givenness : SpeakerType

The analysis returned a marginally significant two-tailed p -value of 0.0505 for the interaction Givenness: SpeakerType; however, since the hypothesis about the Chicano speakers was directional, this result reflects a significant one-tailed p -value of 0.02503. Consider *Figure 1*.

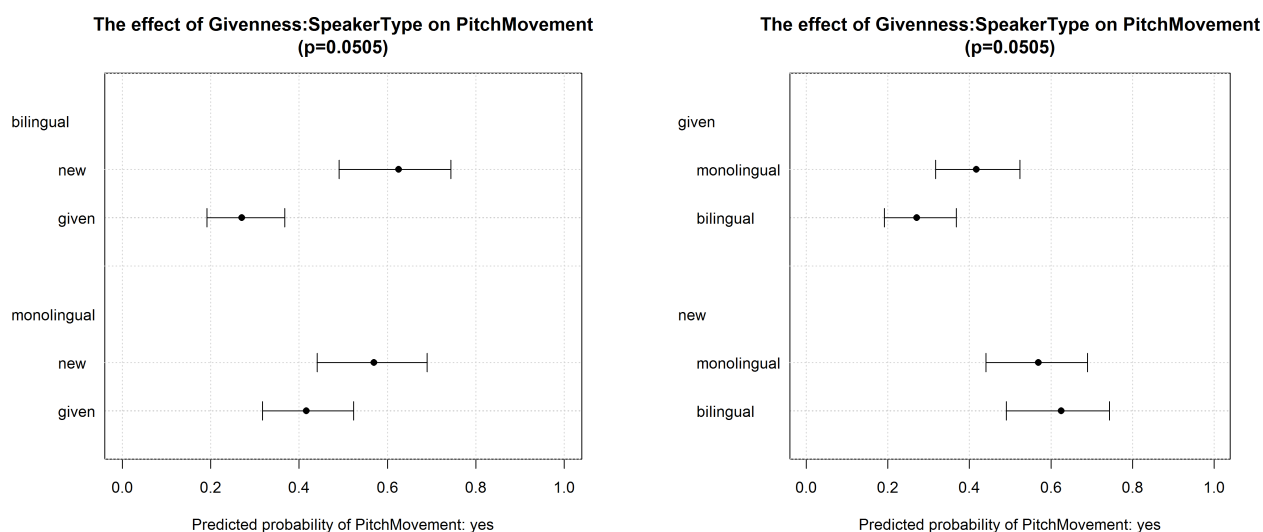


Figure 1. The interaction Givenness : SpeakerType. The x-axis represents the predicted probability that pitch movement occurs across the word. The left and right panels show different perspectives of the same effect.

This interaction shows that the Spanish of both monolingual Mexican speakers and bilingual Chicano speakers uses pitch movement to signal new information. However, while both speaker types use pitch movement for new information, they do not use pitch movement in the same way. Specifically, Chicano speakers use pitch movement more often for new information, as compared to the monolingual speakers; notice that while the direction of this trend is the same for both speaker types, there is no overlap between the predictions for new and given information for the bilingual speakers, while there is an overlap for monolingual speakers. This is to say that bilingual speakers have a more plastic (or English-like) use of pitch, which is compatible with this study's hypothesis. This suggests that the English prosodic system influences the Spanish prosodic system of bilingual speakers. Furthermore, monolingual speakers not only use more pitch excursions for new information, they also use less pitch excursions for given information. Thus it appears that the English prosodic system not only influences bilingual Spanish by causing more pitch movement for new information, but also causes flatter pitch contours in given information.

Phrase Final

Whether the word was phrase final or not was a significant predictor of PitchMovement in the model. *Figure 2* shows that both speaker types are more likely to show pitch movement when a word is phrase final.

The effect of PhraseFinal on PitchMovement

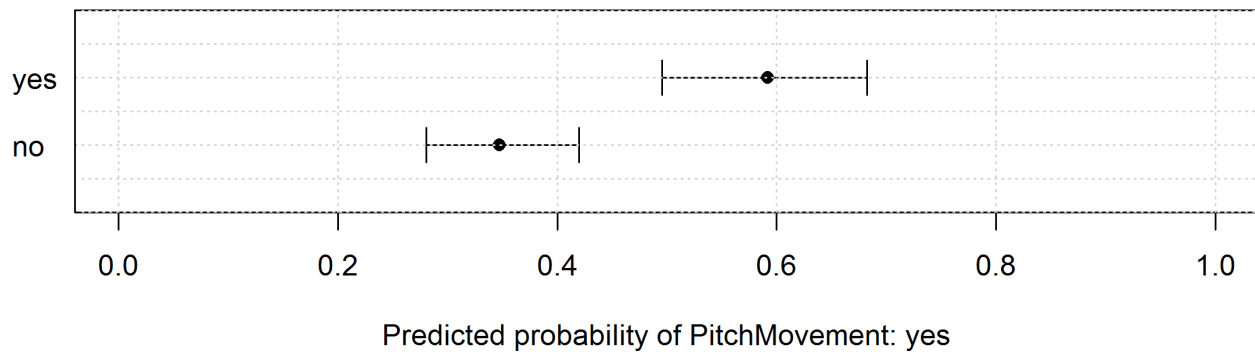


Figure 2. PhraseFinal's effect on pitch movement. The x -axis represents the predicted probability that pitch movement occurs across the word. The phrase final words are represented higher on the graph while those words that are not phrase final are lower.

In addition to signaling new information with pitch excursions, both Chicano and Mexican Spanish speakers use pitch movement for other purposes. Specifically, this effect shows that both speaker types use pitch movement at the boundaries of intonational units; therefore, the position of the analyzed word within the phrase needs to be controlled for. Thus, we can conclude that pitch movement is also used in Spanish (or at least in the Mexican and Chicano dialects of Spanish) as an indicator of phrase boundaries. This corresponds to previous findings for characteristic intonation patterns of Chicano English (Ericson, 2007), suggesting that the same factors that are present in the intonation of Chicano English are likely also present in Chicano Spanish.

LogDuration

The previous two effects have shown that both monolingual Mexican speakers and bilingual Spanish speakers use pitch movement to signal new information and mark the end of phrases. The current effect of LogDuration shows that pitch movement is also correlated with an accompanying increase in the duration of the stressed vowel.

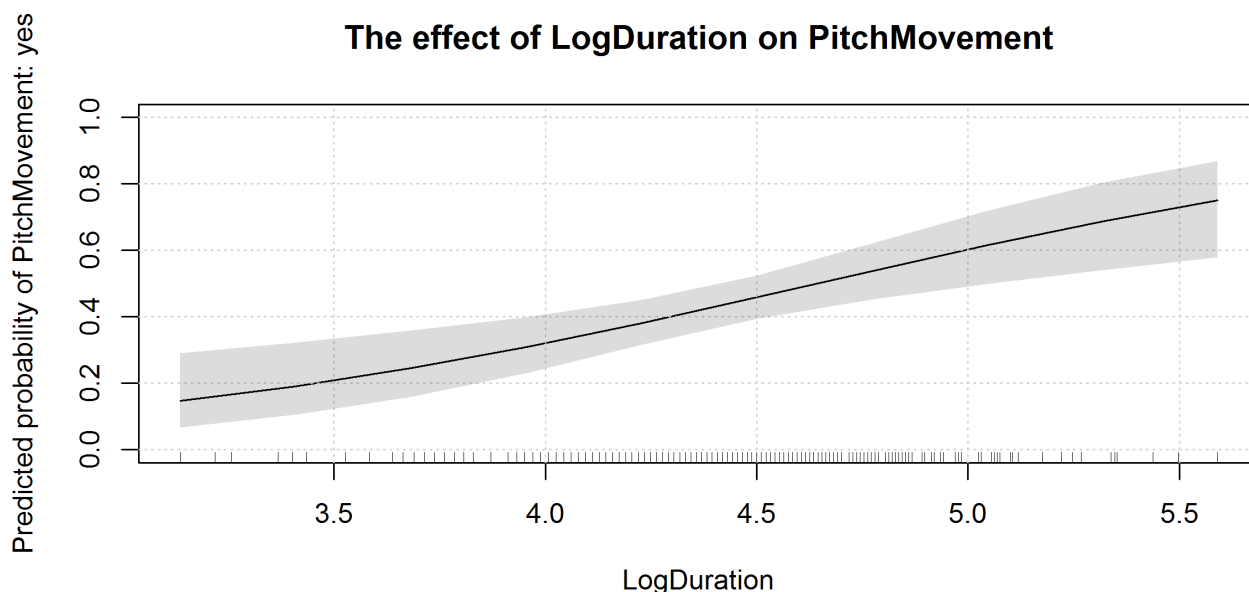


Figure 3. The Log of Vowel Duration’s effect of PitchMovement. Vowel durations are represented on the *x*-axis, with longer durations to the right and the probability of pitch movement is represented on the *y*-axis.

As *Figure 3* shows, more pitch movement is associated with longer durations of tonic vowels. Thus, it is likely that the participants also employ longer durations to indicate new information in conjunction with pitch movement. This is in agreement with previous literature; as mentioned, increased durations work in conjunction with pitch in order to indicate, for instance, lexical stress (e.g. Fry 1958). It should be noted that Fry (1958) also suggested that intensity works in conjunction with these two prosodic cues. However, intensity was not a significant predictor of PitchMovement in our data. This finding could be in tune with previous literature, according to which intensity tends to be a less dependable prosodic cue. In fact, at least in terms of stress, it has been suggested that intensity is a less reliable correlate as compared to segment duration and F0- it is “generally considered a weak cue in the perception of linguistic stress” (Sluijter & van Heuven, 1996, p. 2471).

DISCUSSION

This paper set out to compare the use of pitch by monolingual Mexican speakers and bilingual Chicano speakers in signaling new information. The data suggest that:

- a) pitch movement is used to indicate new information;
- b) pitch movement is also used to mark the end of phrases;
- c) speakers increase vowel duration in conjunction with pitch movement;

Our three research questions are therefore addressed in the following way, As for question 1, we found evidence of dialectal variation between monolingual Mexican speakers and bilingual Chicano speakers in their use of prosody to signal new information using a corpus approach

rather than an experimental one. In terms of a) above, it is clear that heritage speakers use pitch differently than monolingual speakers. Most specifically, and addressing question 2, they use pitch excursions more often to signal new information, and therefore encode new information in a more English-like manner, as compared to monolingual speakers. Question 3 was about what prosodic variables and/or their interactions are relevant in the encoding of information structure by Spanish speakers in general. For this question, a) to c) above indicate that duration and pitch movement are correlates of prosody indicating new information for these Spanish speakers; on the other hand, it appears that neither Chicano Spanish-English bilinguals nor Mexican monolingual speakers use increased intensity as a prosodic marker of information structure, given that the variable Intensity was not a significant predictor in the model. Although heritage speakers are more akin to native speakers than L2 learners in many aspects of their prosody (e.g. Harris & Gries, 2011; Miglio, 2011), our data show that their use of English still affects some aspects of their intonation.

In terms of the applications of this study to language teaching, empirical evidence that differences exist in the intonational systems of two closely related dialects of the same language proves important. The fact that it is the influence of English that apparently causes this dialectal variation is notable, especially given the high number of English-speaking students of Spanish as a second language. For L2 (English speaking) learners of Spanish, we should make them aware that word order can be used to indicate focus in Spanish, and is usually manipulated for that purpose by native speakers, as English speakers are likely unaware of this. We should teach information structure to both L2 and heritage learners, who are likely unaware that speakers use various strategies in order to differentiate between new and given information, as well as teaching intonation in other contexts, such as interrogatives sentences, emphasis, or contrastive focus. Furthermore these data are informative of dialectal differences between L1 and heritage Spanish, which lead to a better understanding of heritage Spanish (for language learners and teachers alike). They serve as an example of a unique feature of Chicano Spanish, which should be represented as a valid dialect of Spanish, rather than a substandard hybrid of Spanish and English. In any case, these data constitute an interesting example of the effect that the prosodic system of one language can have on that of another in bilingual speakers.

ACKNOWLEDGMENTS

The recording of the Mexican participants was made possible with a grant from UC Mexus. Our thanks to Dr. Chantal Melis for facilitating access to the participants and the facilities where they were recorded.

ABOUT THE AUTHORS

Michael Harris (mjh36@juno.com) holds a PhD in Hispanic Languages and Literatures from the Department of Spanish and Portuguese, UCSB. His primary research interest is speech rhythms of Spanish, Portuguese, and English. He believes that multifactorial, computational, and corpus methodologies are effective in addressing linguistic queries and seeks to apply novel and innovative approaches in his research. He has published and/or presented on subjects including speech rhythms, the accusative-oblique alternation in Spanish clitics, and the Portuguese inflected infinitive. Current projects include an empirical study of sociolinguistic attitudes

towards the use of Spanish in the US and a study of dialectal variation in Spanish vowels, as well as ongoing work on speech rhythms of Iberian languages.

Viola G. Miglio (miglio@spanport.ucsb.edu) is currently Associate Professor of Iberian Linguistics and Barandiaran Endowed Chair of Basque Studies in the Spanish and Portuguese Department at the University of California, Santa Barbara (UCSB). Originally a Germanic philologist (MAs from Edinburgh (1989) and Bologna (1993)), she then turned to phonetics/phonology (Ph.D. from the University of Maryland (1999)). She taught English and Romance linguistics at the University of Iceland (1997-2002), before accepting the position at UCSB (since 2002). She works on linguistic rights, translation, Romance languages and Icelandic, and has published articles on Hispanic linguistics, a book on *Interactions between Markedness and Faithfulness Constraints in Vowel Systems* (Routledge 2005, 2012), and recently co-edited a book on *The Protection of Cultural Diversity* (with Xabier Irujo, Center for Basque Studies/University of Nevada Press, 2014).

Stefan Th. Gries (stgries@linguistics.ucsb.edu) is Professor of Linguistics at the University of California, Santa Barbara. Methodologically, he is a quantitative corpus linguist at the intersection of corpus linguistics, cognitive linguistics, and computational linguistics, who uses a variety of different statistical methods to investigate linguistic topics in morphophonology, syntax, the syntax-lexis interface, and semantics as well as first and second/foreign language acquisition. Theoretically, he is a cognitively-oriented usage-based linguist (with an interest in Construction Grammar) in the wider sense of seeking explanations in terms of cognitive or psycholinguistic processes.

REFERENCES

- Arvaniti, A. & G. Garding. (2007). Dialectal variation in the rising accents of American English. In J. Cole & J. H. Hualde (eds), *Papers in laboratory phonology* (9). 547-576. Berlin & New York: Mouton de Gruyter.
- Boersma, P. & D. Weenink. (2010). Praat: doing phonetics by computer (Version 5.1.29 [Computer program]. Retrieved June 21, 2009, from <<http://www.praat.org/>>.
- Butragueño, P. (2004). Configuraciones circunflejas en la entonación del español mexicano. In *Revista de Filología Española*, 84(2). 347-373.
- Cruttenden, A. (1993). The deaccenting and reaccenting of repeated lexical items. *ESCA Workshop on Prosody 1993, Working Papers 41*, Dept. Linguistics, University of Lund, Sweden.
- Deterding, D. (2001). The measurement of rhythm: a comparison of Singapore English and British English. *Journal of Phonetics*, 29(2). 217-230.
- Ericson, H. (2007). An intonational analysis of Mexican American English in comparison to anglo American English. Thesis submitted to the Graduate Faculty of North Carolina State University in partial fulfillment of the requirements for the Degree of Master of Arts.
- Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech*, 1(2), 120-152.
- Harris, M. & S. Gries. (2011). Measures of speech rhythms and the role of corpus-based word

- frequency: a multifactorial comparison of Spanish(-English) speakers. 2011. *International Journal of English Studies*, 11(2). 1-22.
- Miglio, V. (2011). "El doble rostro de los bilingües – las vocales en el español chicano y mexicano". The Two Faces of Fiction/14th Annual Colloquium on Mexican Literature, UCSB, Nov. 3-5, 2011.
- Nazzi, T., J. Berteconi, J. Mehler. (1998). Language discrimination by newborns: Towards understanding the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, 24(3). 756-766.
- Prieto, P. & Roseano, P. (coords.) (2010). Transcription of Intonation of the Spanish Language. Lincom Europa: München.
- R Development Core Team. (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <<http://www.R-project.org/>>.
- Sluijter, A.M. & V.J. van Heuven. (1996). Spectral balance as an acoustic correlate of linguistic stress. *The Journal of the Acoustical Society of America*, 11. 2471-2485.
- Vallduví, E. (1992). *The informational component*. Garland Publishers: New York/London.
- Zubizarreta, M.L. & E. Nava. (2010). Encoding discourse-based meaning: prosody vs. syntax. Implications for Second Language Acquisition. *Lingua*, 121(4). 652-669.

LEARNER PREFERENCE AND THE LEARNING OF JAPANESE RHYTHM

Naoko Kinoshita, Waseda University

This paper reports research which examined the effects of introducing different Japanese rhythm instruction techniques to second language learners. Any successful teaching method needs to take individual variation in both levels of attainment and the learning preferences of the learners into account. Multiple instruction techniques were introduced during Japanese pronunciation classes, which were 1) rhythm marking, 2) clapping, 3) grouping rhythmic patterns, 4) haiku, 5) pronouncing to a beat, 6) visual acoustic analysis (Praat), and 7) shadowing. A questionnaire asking the students their preferred instruction technique administered at the conclusion of the class demonstrated that learners varied in both their learning method preference, and the number of methods they found useful. Of the 25 participants, 16 preferred a single method and the remainder preferred multiple techniques. Using Praat software for visual acoustic analysis and shadowing were the most popular (8 selections each), followed by clapping (7 selections). Participants significantly improved their ability to perceive rhythmic minimal pairs when compared to a pre-test of the same words. They were also able to transfer their improved understanding to a new set of minimal pairs.

INTRODUCTION

This paper examines the effectiveness of second language pronunciation instruction for learning Japanese rhythm. While the original classification system of syllable-timed, stress-timed (Abercrombie, 1967), and mora-timed (Ladefoged, 1975) languages has been discredited (Ramus et al. 1999, Grabe & Low, 2002), rhythm remains an important part of supra-segmental pronunciation.

Although there are a number of acoustic correlates of rhythm (Dauer, 1987), Japanese rhythm is largely determined by the structure of the syllable, and the special morae. The special morae consist of long vowels and diphthongs (*chouon*), geminate consonants (*sokuon*), and syllabic nasal consonants (*hatsuon*). Syllables containing special mora tend towards a similar length of enunciation which is roughly 1.6 times (citation) the length of standard Japanese syllables. Toki and Murata (1989) have summarized the structure of Japanese into standard (short) syllables, and moraic (long) syllables, which is summarized in Table 1. Here vowels are notated as V, semi-vowels as SV, consonants as C and special mora as + (N, Q, R, V, V').

Rhythm is an important part of second language learner speech. The goodness of a learners' rhythm influences native speaker evaluations of comprehensibility and quality (e.g., Derwing & Munro, 1997; Munro, 1995; and Kinoshita & Sheppard, 2011). In spite of this, research investigating the acquisition and teaching of prosody, in general, and rhythm in particular, is sparse (Trouvain & Gut, 2007). In addition, education practice tends to be bias towards one or two techniques (Kinoshita, 2010, 2011).

Table 1

Japanese Syllable Structure

Syllable	Syllable Structure	No. Mora	Example
Standard (Short)	V	1	胃(<i>i</i>), 絵(<i>e</i>)
	CV	1	木(<i>ki</i>), 目(<i>me</i>)
	SVV	1	矢(<i>ya</i>), 輪(<i>wa</i>)
	CSVV	1	書(<i>syo</i>), 茶(<i>cha</i>)
Long (Moraic)	V+ 【N, Q, R, V, V'】	2	恩(<i>on</i>), 運(<i>un</i>)
	CV+ 【N, Q, R, V, V'】	2	金(<i>kin</i>), 禅(<i>zen</i>)
	SVV+ 【N, Q, R, V, V'】	2	ヨ ッ(<i>yo+Q</i>)
	CSVV+ 【N, Q, R, V, V'】	2	今日(<i>kyo+R</i>)
	CSVV+ 【N, Q, R, V, V'】	2	ヒ ヨ イ(<i>hyoi</i>)

Rhythm Acquisition and Instruction

In order to improve the teaching of rhythm in Japanese (and other languages), it can be argued that methodology needs to be informed by understanding of the natural processes of rhythm acquisition. The following is a summary of the findings to date. 1) As with other aspects of pronunciation, it appears that second language rhythm is learned from a first language starting point. In a three year longitudinal study of learners of Japanese as a second language, Kinoshita (2009, 2010) found that they produced rhythm metric scores closer to those of their first language than the target language. 2) Another important finding is learners often failed to develop their rhythm (Kinoshita, 2010, 2011). This is supported by Toda (2003) who suggested that the rhythm and timing of the special mora is particularly difficult to learn. 3) A further point is that rather than a universal rhythm acquisition process, where all learners gradually develop towards native speaker norms, it appears that learners follow different paths. Nakagawa et al. (2008) determined that learners acquire Japanese rhythm more effectively when using a method which better matched their preferred perceptual learning style (Reid, 1987, 1992). A perceptual learning style is the preferred perceptual channel to process information through the working memory (Baddeley, 2000). According to Bernard (1999), working memory has three perceptual processing channels, visio-spatial, auditory, and body state. Individual preference and experience using these processing channels lead to preferred perceptual learning styles. 4) There is also some evidence from the literature that the acquisition of rhythm is similar to skill acquisition. Anderson (1982) proposed that proficient competence in a skill is represented by unconscious procedural knowledge. Kinoshita (2013) demonstrated that second language learner rhythm performance is most likely based on underlying rules. 5) This procedural knowledge is

developed through the power law of practice (Newell & Rosenbloom, 1981), where learners become increasingly more efficient with practice. Practice can be assisted by declarative rule based knowledge, which through conscious attentional systems, can assist the development of the procedural knowledge. Yanagisawa et al. (2013) provided some support for this when they found that learning approaches which enabled the creation of explicit representations of rhythm were effective in second language rhythm acquisition.

It is clear that in order for many second language learners to develop their perception and production of rhythm, it needs to be explicitly taught. Secondly, as rhythm is a skill, rules need to be provided followed by the opportunity to internalize them. This should be followed by ample opportunity to practice rhythm through application of the rule in a controlled environment. Finally, instruction would likely benefit from taking the individual learning processes into account. This could be done is by introducing a number of different instructional techniques focusing on different perceptual learning styles during the course.

Research Questions

In this paper, a course which used this approach to rhythm instruction was evaluated using an action research method. The research aimed to determine the effectiveness of a multi-instructional technique approach. The following research questions were investigated.

- RQ1: After a semester-long pronunciation course in which students were exposed to seven instruction techniques to learn Japanese rhythm, which technique(s) did learners prefer?
- RQ2a: Did the second language learners of Japanese improve their perception of Japanese rhythm after this instructional intervention?
- RQ2b: Did improvement demonstrated on pre-post test transfer to the perception of rhythm differences on novel lexical items?

METHOD

Participants

The participants in this study were 25 university students enrolled in a lower intermediate-level Japanese pronunciation course. The majority of the participants had studied Japanese in their home institutions, and had arrived in Japan between 3 weeks prior to the beginning of the course. Enrolment and level was self-determined, leading to variety of ability levels in the class. They came from a variety of different countries and spoke a large number of first languages (China, 6; Hong Kong, 3; USA, 3; England, 3; Ireland, 2; Indonesia, 2; Korea, 2; Russia, 2; and Taiwan and Uganda, 1 each). The students agreed to allow their data to be used in this study.

The course was run over a 15 90-minute classes, once a week. The goals of the course were as follow: 1) understand the rules of Japanese rhythm, accent, and intonation, 2) understand the characteristics of their own pronunciation, 3) be able to apply the rules of pronunciation during speech, and 4) determine a more appropriate rhythm instruction technique.

Procedure

At the beginning of the rhythm section of the course, the participants were required to complete a pre-test, which represented their perceptual rhythmic knowledge. They were then introduced to the different pronunciation instruction techniques, described in detail below, and experienced learning pronunciation through each of them. Each method was introduced over about 40 minutes, half of a weekly class. After the four weeks, the participants then completed a post-test of rhythm which consisted of the same six items as the pretest. A second test (post-test 2), which contained new items was administered following the first post-test to check if there had been transfer of knowledge. In the final class of the semester, the participants filled out a learning preferences questionnaire and an agreement to participate in the study. The different instruction techniques are described in detail below and summarized in Table 2:

Table 2

A Summary of Japanese Rhythm Instruction Techniques

Instruction Technique	Visual	Auditory	Kinesthetic	Explicit Training, Explanation (Minutes)	Time on Task (Minutes)
Rhythmic Marking	X		X	10	20
Clapping			X	10	20
Pattern Grouping		X		10	20
Metronome Haiku		X		10	25
Auditory Beat		X		10	25
Acoustic Analysis	X			15	25
Shadowing		X		15	25

Rhythmic marking

Rhythmic marking is a method where the teacher begins by explaining the difference between simple and long syllables. Next, the learners are asked to mark the long and the short syllables shown in Figure 1. Then students trace the symbols with their fingers as they listen to the teacher and say the words aloud. As the learners build up these visual, aural, and tactile representations, the complexity is increased first by placing the words into carrier sentences, and then by placing them into more authentic communicative utterances.

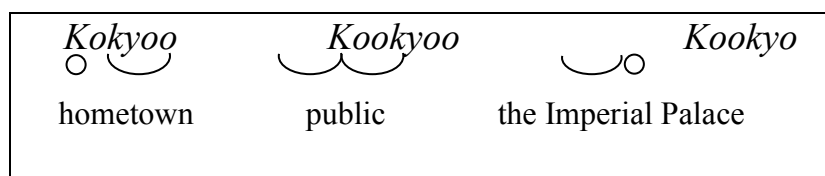


Figure 1. Rhythm marking annotation.

The objective of rhythmic marking is to assist the learners to created explicit visual and kinesthetic representations of the target rhythmic structure. It is hoped that the learners are able to transfer these representations to implicit auditory and oral representations through multiple practice sessions and conscious application of the process developed above

Clapping

As above, this method starts with an explicit explanation of the difference between long and short syllables. The students then mimic the teacher as he/she reads words out loud while clapping once for a short syllable, and twice for a long syllable (see Figure 2). Similarly to the process in Rhythm Marking, the utterances increase gradually in complexity; from single words, to simple sentences, to multiple clause sentences.

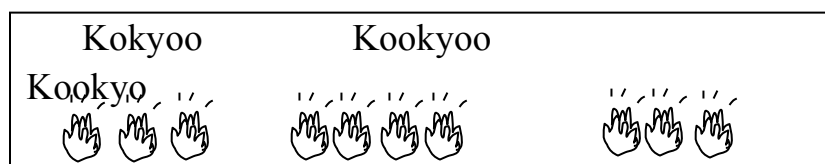


Figure 2. Visualization of clapping.

This method is thought to contribute to learning by assisting learners to create a kinesthetic representation of rhythm. Once this representation is established, the learners are able to base their repetitive practice on conscious application of it. This is different to Rhythm Marking in that it does not have a visual component to the representation.

Rhythmic Pattern Grouping

This approach requires learners to group lists of words currently in their lexicon into their respective rhythm patterns. A 1-2 pattern consists of a word made up of a short and a long syllable. A 1-2-1 pattern word is one with a short, a long, and a short syllable. Examples are provided in Figure 3. (The rhythmic marks are added to denote short and long syllables in the figure, and are not actually part of the teaching technique). Once the lists have been created, the learners should read the words aloud repetitively, vocalizing the pattern



Figure 3. An example of Rhythmic Pattern Grouping.

This method assists learners to create a conscious auditory representation of the different rhythmic structures, and provides the opportunity to internalize them through repeated practice. It is different from the above method as it does not provide kinesthetic support. Nor does it provide the opportunity to generalize representations to sentences. The effectiveness of this method is reported and discussed in Kinoshita (2013).

Metronome Haiku

This method introduces Japanese haiku, which are structured around mora-rhythm; five morae in the first line, seven in the second, and five in the final line. First, the teacher reads a haiku out loud, as a metronome beats time. Two morae are pronounced with each metronome click, and the pause between each line takes half a beat. The learners read in time with the teacher and metronome. Following this, learners create their own haiku and read them aloud in time to the metronome.

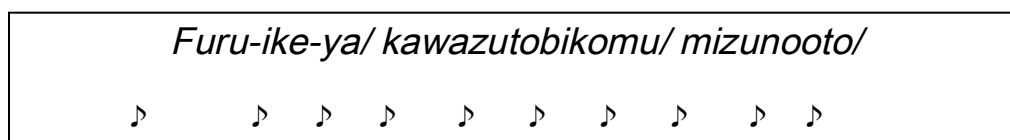


Figure 4. Metronome Haiku beat placement.

The aim is that learners will be able to create auditory representations of mora-timing and a sense of isochronic timing through this technique. Although this method does not include a large amount of practice, it is hoped that the learner is able to apply the representations they learn through this technique when practicing using other methods, and through communication.

Auditory Beat

In Auditory Beat, a similar technique to Metronome Haiku, students read several short sentences with similar rhythmic structure aloud to a musical beat. This method is much like jazz chants which are often used for teaching English stress and timing to children (Graham & Veltfort, 1979). Auditory Beat assists acquisition in similar ways to Metronome Haiku. However, the rhythmic structures that can be used are not restricted by the structure of haiku. This method can also provide students with the opportunity to practice which will assist in the internalization and automatization of the conscious representation.

Visual Acoustic Analysis

Learners using Visual Acoustic Analysis as a method compare their own pronunciation with that of a model. They are first required to learn to use acoustic analysis software such as Praat (Boersma & Weenick, 2014). After learning how to record, upload, and analyze length in their utterances, they compare their own pronunciation with that of a pre-recorded native speaker (Figure 5). The learner then re-records their utterance in an attempt to pronounce the word/phrase/sentence with a sound pattern closer to the native speaker model.

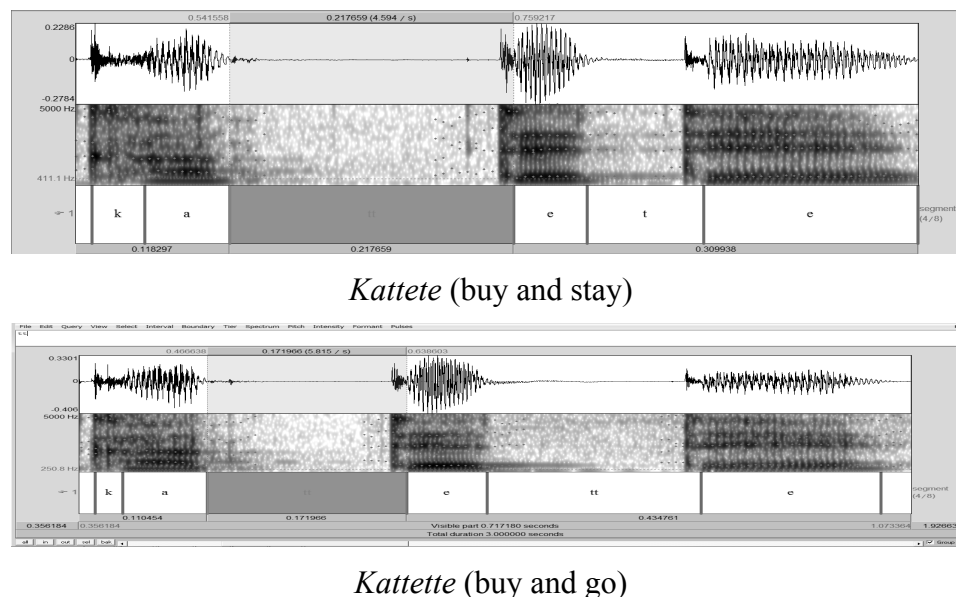


Figure 5. Comparing *Kattete* and *Kattette* with Praat (2014).

Visual Acoustic Analysis offers the opportunity for visual feedback on production. Students are given the ability to compare their production with a model and are then enabled to make adjustments to their rhythmic production. This assists in building an accurate, conscious representation of the target sound patterns. As the comparison can be done at any time, it also enables the learners to monitor their performance outside the limited capacity of their working memory.

Shadowing

Shadowing is the final pronunciation learning method introduced during the class described in this paper. Using this method, the learners first watched a drama, focusing on comprehending the content. Once the story was understood, the drama script was provided, and the drama played once again, while learners listened to the drama, and read the script aloud in time with it. Once they had shadowed the drama, the learners then recorded their production and checked their own pronunciation and rhythm by comparing it to the drama script.

The mechanism of acquisition through shadowing is discussed in Kadota (2007). This technique differs from the others in that it attempts to build an implicit representation of the target pronunciation structures directly through a large amount of practice. The method relies on encoding a segment of the auditory information from the drama in the auditory loop of the working memory, matching it with its meaning through the pre-shadowing viewing of the drama, and the coinstantaneous reading of the text, and then forming one's own representation through conscious rehearsal and monitoring performance by comparison with the representation encoded in the working memory (see Atkinson & Shiffrin, 1971). This method assumes that through the process of shadowing, implicit representations are created, which can be generalized to new contexts.

Analysis

Variables measured in order to answer the above research questions were learning method preference and achievement. Learning method preference was measured by questionnaire. It required the participants to select (with a check-mark) the methods which they believed matched their own way of learning or motivated them most. Multiple responses were allowed, and learners were given the opportunity to add additional methods which they thought effective, but were not covered in the course.

Achievement was measured by performance on a rhythm test. The test consisted of a 6-item listening test which required the participants to listen to and identify the correct rhythmic pattern from a number of choices representing different rhythmic patterns. For example: kokyo (hometown), kookyoo (public), kookyoo (the Imperial Palace). A second version of the test was created to determine if what was learned was transferable to a new context. The test items for this perceptual rhythm discrimination test are provided in Appendix A.

RESULTS

RQ1: After a semester-long pronunciation course in which students were exposed to seven instruction techniques to learn Japanese rhythm, which technique(s) did learners prefer?

The first research question set out to determine what method learners prefer. Figure 6 shows the number of students who selected each of the learning methods as their preference. 16 participants selected just one method, the remaining participants choose an average of three methods. Two participants wrote 'song' as their preferred method of learning, even though it was not one of the methods covered in class. Of the techniques which were covered in class, Visual Acoustic

Analysis (Praat) (8), Shadowing (8), and clapping (7) were the most popular. Grouping and haiku were the least popular of the methods. The remainder, marking and beat were both selected by about 16% of the participants (4).

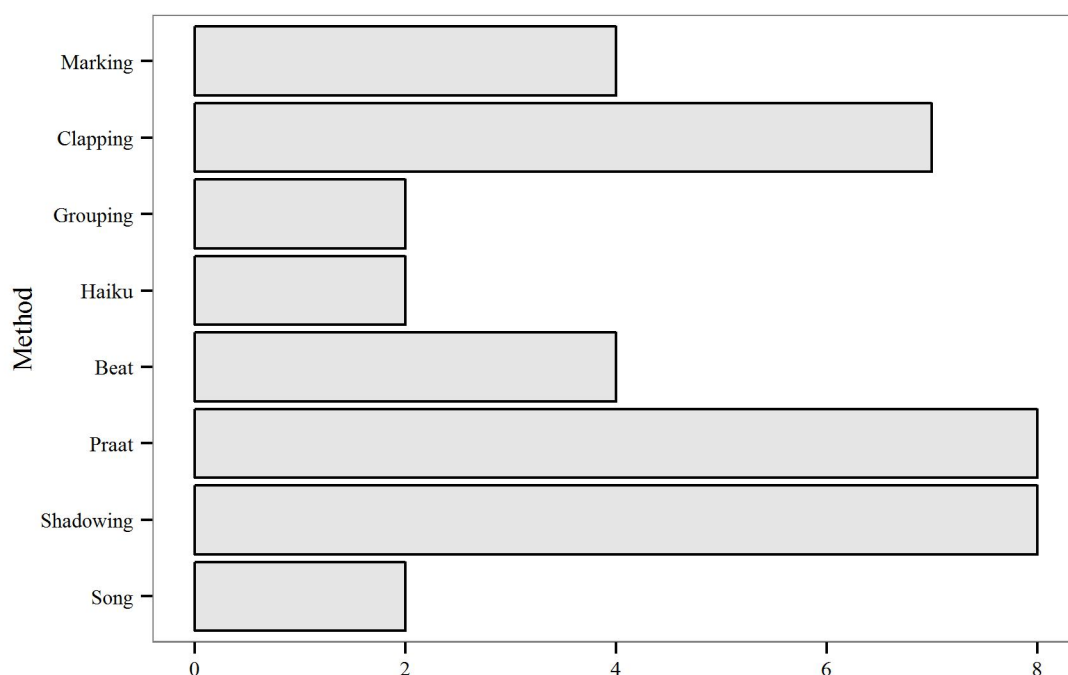


Figure 6. Participant Preferences for Learning Japanese Pronunciation (Rhythm).

The results demonstrate that there is a wide range of individual difference in both learning preference and the number of preferred methods. Just over half preferred a single method while the remainder preferred a larger number of methods. It is also apparent that the participants were not biased towards a particular learning style. The three most popular techniques were based on different perceptual learning styles: Visual Acoustic analysis (8) can be considered to be primarily visual, Shadowing (8) relies largely on the auditory channel, and Clapping (7) is primarily an auditory and kinesthetic method.

RQ2a: Did the second language learners of Japanese improve their perception of Japanese rhythm after this instructional intervention?

RQ2b: Did improvement demonstrated on pre-post test transfer to the perception of rhythm differences on novel lexical items?

The second research questions examined the overall effectiveness of providing the learners with a variety of instruction techniques. Figure 7 depicts the change in the average Japanese rhythm scores over the pre-test and the two post-tests. The graph shows a linear increase from the pre-

test to post-test one then to post-test two. There is also a corresponding decrease in the test score standard deviation. A final point that can be made here is that it is likely that the second post-test is demonstrating a ceiling effect as the maximum score of 6 is within one standard deviation of the average. A Repeated Measures ANOVA (corrected for sphericity violations) ($F(2, 48) = 11.172, p < .001, \eta^2 = 0.19$) supported this result. Post-hoc tests (Bonferroni correction) demonstrated that both post-test 1 ($p < .05$) and post-test 2 ($p < .05$) were significantly better than the pre-test.

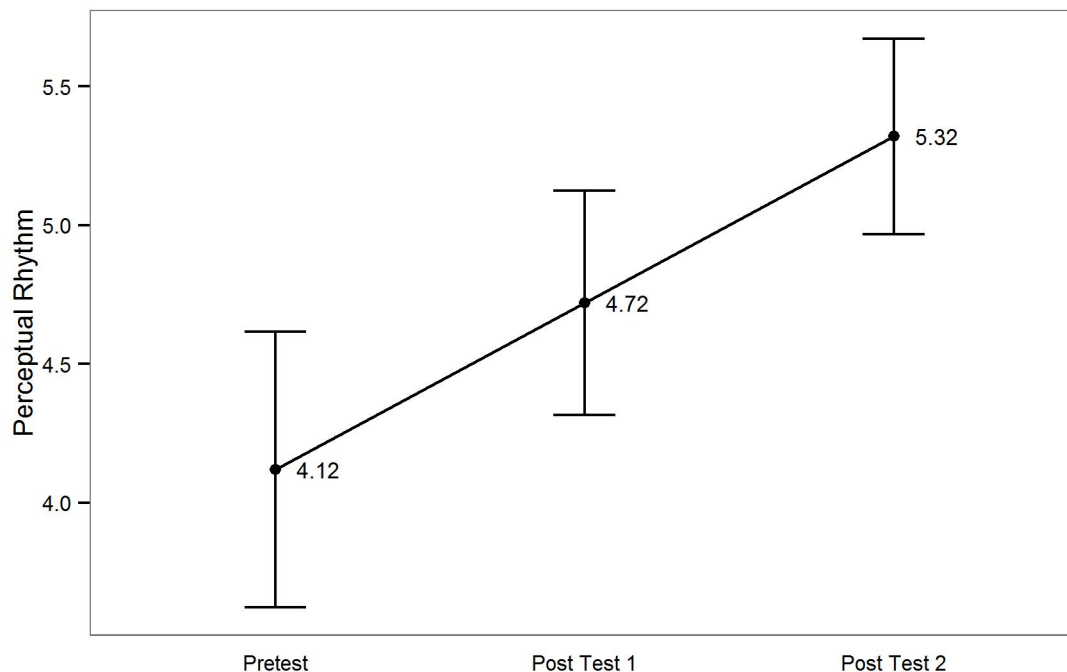


Figure 7. Pre- and post-test means of the rhythm test (error bars represent standard error).

These tests demonstrated that the learners were able to improve their ability to perceive differences in Japanese rhythm. This ability is also transferred to a new set of test items. This transfer is important because of the study's design. As it was unethical to create a control group, some way to demonstrate that the improvement was not only due to taking the same test twice was needed. A variation on time-series designs (i.e. Mellow et al., 1996) was used here to attempt to overcome this weakness. It was hoped that the performance on the new test items used in post-test 2 would reflect overall improvement, and mitigate the practice effect that may have been present as a result of doing the pre-test twice. There is an underlying assumption here that if the items in the two test versions are equivalent and the participants had completed the second test version as a pre-test, they would have had similar scores to that of the actual pre-test.

DISCUSSION

The main findings demonstrated that 1) learners have different preferences when it comes to learning, and that using varied instruction techniques results in improvement in Japanese language rhythm. Taken together, these results provide some limited support for a teaching approach which uses multiple instructional approaches. This is supported by Alvermann (2002) in reading development, Francisco et al. (1998) in science education, and Ainsworth (2006) in general education. Providing an opportunity to experience different Japanese rhythm instruction techniques most likely enabled the learners to find, and learn through method/s which best matched their own perceptual learning styles. This may have resulted in the improvement of Japanese rhythm which was measured by perceptual discrimination.

The results of this study suggest that future classes should continue to present a number of instruction techniques, and provide multiple opportunity to practice, rather than just teaching the rules of rhythm explicitly. Perhaps more time can be spent on the more popular methods of shadowing, clapping, and Praat. It is also important that time is spent actually using each method in class and at home, so that the learners will have enough experience to know which of the methods most suits their learning.

Future research needs to clarify if using preferred methods of learning leads to more improvement in rhythmic ability when compared to using less-preferred methods and would benefit from the inclusion of a control group. In addition a follow up study should be conducted to determine if students continue learning using the methods they selected, and if their pronunciation continues to improve outside of the classroom environment. Other limitations of this study stemmed from the self-report nature of the learners preferred instruction technique and the fact that there was no direct measurement of the perceptual learning style of the learner.

ACKNOWLEDGMENTS

This paper is a part of the outcome of research performed under a Waseda University Grant for Special Research Project (Project number: 2014S-162). I would also like to acknowledge the assistance of the Center from Japanese Language, Waseda University for ethical approval and express my special appreciation to the participants of my study. Thank you also to the anonymous reviewer for his/her suggestions and comments.

ABOUT THE AUTHOR

Naoko Kinoshita (kinosyta@gmail.com) received her Ph.D. in 2010 from Waseda University. She has taught Japanese as a Foreign Language in Busan University of Foreign Studies, Korea, and as a second language in Meikai University in Japan. She is currently working at the Japanese Language Education Center at Waseda University in Japan as an Associate Professor. Her research interests focus on the education and acquisition of second language pronunciation. She has published research in the measurement, the acquisition, and education of second language rhythm acquisition. She is currently working on three projects. The first has the goal of creating an on-demand learning system where students can continue learning Japanese pronunciation autonomously. The second is investigating the role of individual word syllabic structure on the development of generalized rhythm representations. The third deals with the development of critical thinking skills in Japanese universities.

REFERENCES

- Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction*, 16(3), 183–198.
- Alvermann, D. E. (2002). Effective literacy instruction for adolescents. *Journal of Literacy Research*, 34(2), 189-208.
- Atkinson, R. C., & Shiffrin, R. M. (1971). *The control processes of short-term memory*. Technical Report 173. Palo Alto: Stanford University.
- Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, 89, 369-406.
- Baddeley, A.D. (2000) Working memory and language processing. In B. E. Dimitrova & K. Hylltenstam (Eds.) *Language processing and simultaneous interpreting: Interdisciplinary perspectives*. Amsterdam/Philadelphia: Benjamins Translation Library.
- Boersma, Paul & Weenink, David (2014). Praat: doing phonetics by computer [Computer program]. Version 5.4.01, retrieved 9 November 2014 from <http://www.praat.org/>
- Dauer, R. M. (1987). Phonetic and phonological components of language rhythm, *Proceedings of the XIth ICPHS* (pp. 447-450). Tallinn: Estonia.
- Derwing, T. M., & Munro, M. J., (1997). Accent, intelligibility, and comprehensibility. *Studies in Second Language Acquisition*, 19 (1), 1-16.
- Francisco, J.S., Nicoll, G., & Trautmann, M. (1998). Integrating multiple teaching methods into a general chemistry classroom. *Journal of Chemistry Education*, 75 (2), 210-213.
- Grabe, E., & Low, E. E. (2002). Durational variability in speech and the rhythm class hypothesis. *Papers in Laboratory Phonology*, 7, 515-546
- Graham, C., & Veltfort, A. (1979). *Jazz chants for children: Rhythms of American English through chants, songs and poems*. Oxford: Oxford University Press.
- Kadota, S. (2007). *Shadowing to ondoku no kagaku* (Shadowing and the science of reading aloud). Tokyo: Kosumopia K.K
- Kinoshita, N. (2011) *Nihon-go norizumu shutoku kennkyuu* (The acquisition of Japanese rhythm), Tokyo: Waseda University Press.
- Kinoshita, N., & Sheppard, C. (2011). Validating acoustic measures of speech rhythm for second language acquisition. *Proceedings of the ICPHS XVII* (pp. 1086-1089). Hong Kong.
- Kinoshita, N. (2013). The acquisition of Japanese special mora rhythm by Chinese learners of Japanese: Comparing known words, unknown words, and non-words. *Meikai Japanese Language Journal* 19, 35-44.
- Ladefoged, P. (1975). *A course in phonetics* (1st ed.) Orlando: Harcourt Brace.
- Mellow, J. D., Reeder, K., & Foster, E. (1996). Using time-series research designs to investigate the effects of instruction on SLA. *Studies in Second Language Acquisition*, 18, 325-350.
- Munro, M. J., (1995). Nonsegmental factors in foreign accent. *Studies in Second Language Acquisition*, 17, 17-34.

- Nakagawa, C., Sheppard, C., & Kinoshita, N. (2008). A comparison of the learning methods of more and less successful learners of Japanese pronunciation, International conference of teaching and learning (p. 55). Multimedia University, Malaysia.
- Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition*, (pp. 1-55). Hillsdale, NJ: Erlbaum.
- Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73(3), 265-292.
- Reid, J. (1987). The learning style preferences of ESL students. *TESOL Quarterly* 21(1), 87-111.
- Reid, J. (1992). A computer text analysis of four cohesion devices in English discourse by native and nonnative writers. *Journal of Second Language Writing*, 1(2), 79-107.
- Reid, J. (1995). *Learning styles in the ESL/EFL classroom*. New York: Heinle and Heinle.
- Toda, T. (2003). Acquisition of special morae in Japanese as a Second Language. *Journal of the Phonetic Society of Japan*, 7(2), 70-83.
- Trouvain, J., & Gut, U. (2007). *Non-native prosody. Phonetic description and teaching practice*. Berlin: Mouton de Gruyter.
- Yanagisawa, E., Kinoshita, N., & Nakamura, N. (2013) Karada no ugoki o mochiita tokusyuhaku shido no kokoromi – chikaku gakusyu sutairu ni cyumokushite. *The Society for Teaching Japanese as a Foreign Language syuki taikai*, 405-406.

APPENDIX

Table 1

Items for the pre/post and post-post perceptual rhythm discrimination test.

No.	Test Items	
	Pre/Post Test	Post-Post Test
1	tori - toori	kako - kakko
2	koki - kokki	ido - idoo
3	mina - minna	kinen - kinnen
4	kyouju - kyoujuu	jocyou - joucyo - joucyou
5	saka - sakka - sakkaa	gaka - gakka
6	ryokou - ryoukou - ryouko	kokyou - koukyo - koukyou

WHEN FRENCH BECOMES TONAL: PROSODIC TRANSFER FROM L1 CANTONESE AND L2 ENGLISH

Jackson L. Lee, University of Chicago

Stephen Matthews, University of Hong Kong

What happens when native speakers of a tone language learn a non-tone language? This paper describes and accounts for L3 French prosody by L1 Hong Kong Cantonese speakers competent in English as L2. The general observed pattern in L3 French prosody is that syllables of French content words (nouns, verbs, adjectives, and so forth) all bear the Cantonese high level tone, whereas syllables of French function words (prepositions, determiners, etc.) have the Cantonese low level tone. This is analyzed as a case of interlanguage transfer, where L1 Cantonese contributes to the observed prosodic features in L3 French via L2 English interlanguage. Beyond the empirical contribution, this paper remarks on pedagogical aspects of second and third language acquisition.

INTRODUCTION

This paper reports observations regarding the prosody of non-native French spoken by native speakers of Hong Kong Cantonese competent in English as a second language, and provides an account for where the observed patterns came from. The overall observed non-native French prosodic pattern is fairly simple in descriptive terms: syllables of French content words (nouns, verbs, adjectives, and so forth) all bear the Cantonese high level tone, whereas syllables of French function words (prepositions, determiners, etc.) have the Cantonese low level tone. Such a prosodic pattern, with a binary tone system consistently differentiating function and content words, is not found in French, Cantonese, or English. We show that such a peculiar non-native French prosody results from language transfer effects. Specifically, our analysis is in terms of interlanguage transfer (Cenoz et al. 2001, Leung 2007), from L2 English interlanguage (English influenced by L1 Cantonese) to L3 French in our case.

In the following, we first explain the importance of documenting such observations in terms of research on prosody and third language acquisition. The necessary linguistic background on Cantonese and French is then provided, in preparation for the description of the non-native French prosodic patterns. We account for the observed patterns in terms of language transfer. The paper concludes with pedagogical remarks.

This paper contributes what would otherwise be hardly known to our empirical database of linguistic varieties. Specifically, we study a form of interlanguage with French being the target language strongly influenced by the learners' native Cantonese and non-native English, and we focus on prosody in the sense of tone and intonation. Beyond the specific languages in question, our case study fills gaps in two research areas: prosody in non-native varieties and third language acquisition, both of which are to be elaborated below. We also discuss the relevant linguistic and prosodic features of Cantonese and French as the necessary background for the description and explanation of the observed prosodic patterns in non-native French.

Prosody in Non-native Varieties

In the literature on second and foreign language acquisition, it is well recognized that prosody, viz. tone and intonation, is among the hardest things to teach and to master well; see Levis (1999) for a review. Paradoxically, this has not quite generated a lot of research on prosody in non-native varieties, in contrasts with the relative abundance of work on morphosyntax and segmental phonology. This can reasonably be attributed to the highly gradient nature of tone and intonation, which makes description and analysis much less straightforward than the much more categorical concepts in such linguistic entities as tense/aspect markers as well as consonants and vowels.

Around half of the world's languages are tone languages (Yip 2002, Maddieson 2013). What happens prosodically (in terms of tone and intonation) when a tone language speaker learns a non-tone language? While we expect effects of language transfer, what factors contribute to the prosodic grammar of the target non-native language? By studying the prosodic features of the non-native French by Cantonese speakers, this paper represents an attempt to contribute to the under-researched area of prosody in non-native varieties.

Third Language Acquisition

Given that there are more non-native speakers of English than native speakers worldwide (Crystal 2010), and that many people who speak English as an L2 learn yet another language -- the third language, there is a growing literature on third language acquisition (Cenoz et al. 2001, Leung 2007). Learning French in Hong Kong constitutes a case of third language acquisition given the socio-historical background of Hong Kong, cf. Yip and Matthews (2012). While the vast majority of individuals born and raised in Hong Kong speak Cantonese as their first language (L1), English is their second language (L2) taught at school since early childhood.¹

There are notable linguistic similarities and differences between French, English, and Cantonese. French is much more similar to English than to Cantonese, in terms of the writing system, vocabulary, morphosyntax, and so forth. Therefore, for L3 French learners with a good command of English in Hong Kong (particularly when language classes are taught with English as the medium of instruction – as was the case for the speakers in our data), it appears intuitively tempting to take advantage of knowledge about English to aid French learning, cf. Kellerman's (1979) psychotypology on learners' perception of similarity among languages). Given such background settings, it is reasonable to consider the possibility that L2 English plays an important role in shaping the prosodic features of L3 French.

Cantonese

¹ Mandarin Chinese is another L2 for the majority of contemporary Hong Kong locals. It is not clear how this might influence non-native French prosody.

² Audio demonstrations for Cantonese tones are available on the book website of Matthews and Yip (2011): <http://www.cuhk.edu.hk/lin/cbrc/CantoneseGrammar/>

Cantonese is a tone language where every syllable has one of the six tones (see Matthews and Yip 2011 for more details):²

The six tones in Cantonese:

Level – high, mid, low

Contour – high rising, low rising, low falling

It has been reported that there is transfer of tonal features from L1 Cantonese to L2 English (Luke 2000, Cheung 2008, Gussenhoven 2012, Yiu 2014). The three level tones in Cantonese are of particular interest here, because they appear to be what is often transferred to the target language by L1 Cantonese learners.

French

In contrast to Cantonese, French is not a tone language. Moreover, unlike languages such as English and Spanish, French has no word stress. The prosody of French, therefore, centers on phrasal and sentential intonation; see Di Cristo (1998), Post (2000), and Gussenhoven (2004) for more detailed descriptions together with analyses in various theoretical frameworks.

For the purposes of this paper, it is sufficient to familiarize ourselves with the intonation pattern of a canonical statement in French. In general, the way French intonation works is that an utterance is syntactically parsed into its constituents (noun phrases, verb phrases, etc). Each constituent then has its own intonational pattern. Typically, it is an overall rising pitch pattern for a non-sentence-final constituent and a falling pattern for a sentence-final constituent. In other words, a common intonation pattern for a statement uttered in a neutral and canonical way has a succession of rising pitch contours, each spanning across a syntactic constituent, until the end of the statement where the pitch contour falls.

Another characteristic of French important for understanding L3 French prosody described in this paper is word order. French is head-initial, similar to English. This means that a phrase which is a syntactic constituent has its head on the left and its complement on the right; for instance, a prepositional phrase has the preposition on the left and its complement noun phrase on the right. It will be shown below that these facts about French are important for accounting for the L3 French prosody by L1 Cantonese speakers.

Data

We recorded classroom spontaneous French as spoken by undergraduates presenting coursework in French language classes at the University of Hong Kong, where the medium of instruction is English. All the undergraduates were native speakers of Hong Kong Cantonese with a good command of English as their L2. They were all third-year undergraduates at the time of being recorded, having taken French language classes at the university for almost 400 class contact hours and having been to France for a summer intensive language program. The data were class presentations where each student spoke for about three minutes. In this preliminary study, we report the most salient L3 French prosodic features by two students, a female and a male. The data were coded in terms of high or low pitch for each syllable. While a more detailed

² Audio demonstrations for Cantonese tones are available on the book website of Matthews and Yip (2011): <http://www.cuhk.edu.hk/lin/cbrc/CantoneseGrammar/>

examination of our dataset is in order, the L3 French prosodic patterns described in this paper appear to be common among the students in our corpus.

L3 FRENCH AS A TONE LANGUAGE

Descriptively, the prosody of L3 French spoken by L1 Cantonese learners follows a simple pattern: all syllables of content words (nouns, verbs, adjectives, etc.) bear the Cantonese high level tone (one of the six tones of Cantonese; see above), whereas syllables of function words (prepositions, determiners, etc.) have the Cantonese low level tone. This effectively makes L3 French a tone language, albeit in a rather unusual sense. There are tone languages such as Cantonese for which the tone of a particular word or syllable is usually unpredictable, as well as others such as Bantu languages for which tone may mark grammatical information such as tense and aspect. However, neither is the case for the L3 French prosody described here: the status of a word as a content or function word correlates with which tone is used.

Using the formalism from autosegmental phonology (Goldsmith 1976, 1990) in the examples below (Figures 1 and 2), we indicate the pitch level of each syllable by using H for high pitch and L for low pitch, with each H or L linked to the respective syllable by a vertical line. The pitch tracks from Praat (Boersma and Weenink 2008) for both examples are also provided. Figure 1 comes from a male speaker, and Figure 2 from a female speaker.

H	H	H	L	H	H	L	H	H	H	H	H	H
aujourd'hui, nous voulons vous présenter quelque chose today, we want you present something												

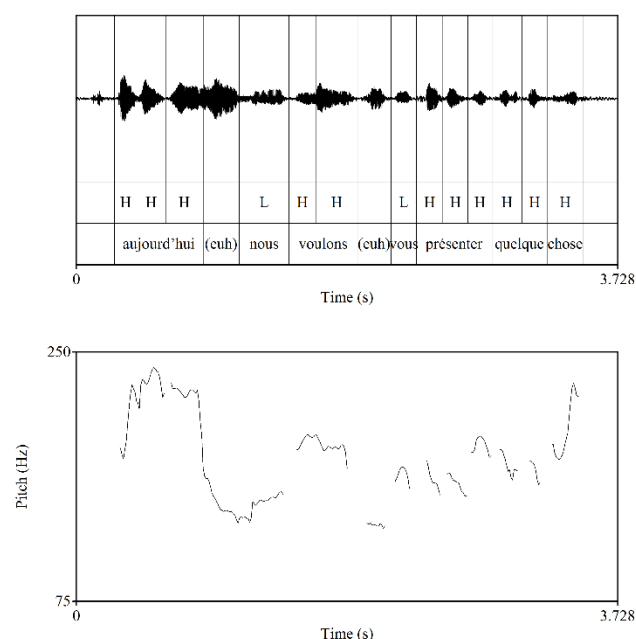


Figure 1. Example of L3 French prosody by a male speaker

H H L L H H H H H H L H
 | | | | | | | | | |
 Paris est la capitale touristique du monde.
 Paris is the capital touristy of-the world

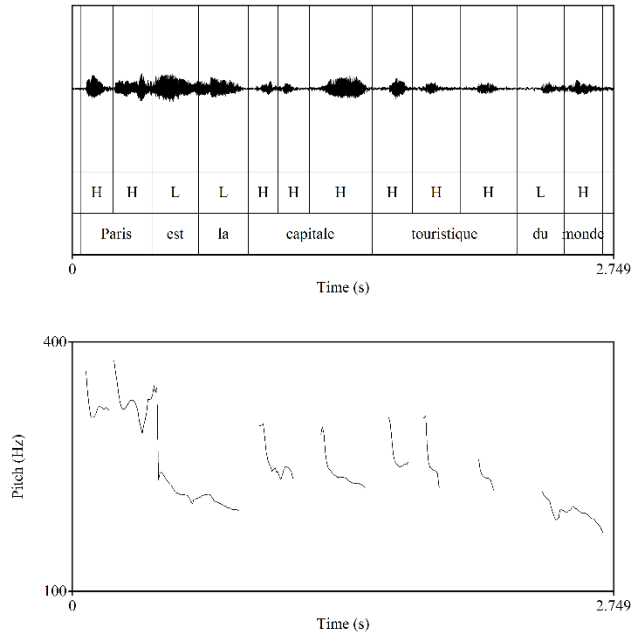


Figure 2. Example of L3 French prosody by a female speaker

The pitch tracks in Figures 1 and 2 appear to suggest that the intonation in both cases is falling in pitch in general, which might cast doubt on whether there are only two discrete level tones being used. To the extent that the pitch patterns are *acoustically* falling, it is well recognized that this phenomenon, known as declination, is universal and physiologically based, due to the fact that pulmonic pressure decreases across an utterance. This means that, for tone languages where pitch is crucial for tone identification, there is presumably a perceptual mechanism that compensates declination across an utterance so that, for example, phonologically identical high tones at two different positions within an utterance are perceived as the same entity despite their acoustic pitch difference. Such a mechanism has been confirmed to exist for Cantonese (Yuen 2007).

There are undoubtedly other prosodic patterns in L3 French by L1 Cantonese speakers that are worth further investigation, e.g., boundary tones as evidenced by the rising pitch contour towards the end in Figure 1 for upcoming new information in the discourse and a falling one in Figure 2 to indicate the end of a statement. Nonetheless, these do not mask the most striking and consistent patterns of Cantonese high level tone used for French content words and Cantonese low level tone for French function words.

ANALYSIS

Where did the L3 French prosodic features reported in the previous section come from? The goal of this section is to provide an account, particularly in terms of interlanguage transfer (Cenoz et

al. 2001, Leung 2007). While French (the target language) is clearly relevant in shaping the learners' French, L2 English (influenced by their L1 Cantonese) plays an indispensable role as well.

Both the intonation and syntax of French contribute to the observed L3 French prosody by Hong Kong Cantonese speakers. As described above, French has no word stress and relies entirely on phrasal and sentential intonation for its prosodic grammar. Non-sentence-final phrases typically have a rising pitch contour. In terms of word order, French is head-initial, which makes a phrase usually begin with a function word and end with a lexical content word.

These facts about the intonation and syntax of French serendipitously produce the following configuration: function words coincide with low pitch, whereas the final syllables of content words bear higher pitch. Despite this, native French prosody does not exhibit what appears in L3 French prosody to be plateaux of high pitch spanning over the entire durations of polysyllabic content words.

Abstracting from important details, such a crude mapping between high/low pitch and content/function words respectively in L1 French matches the observed L3 French prosody by L1 Cantonese speakers where high level tone is used for content words and low level tone for function words. This is by no means the whole story, however. In particular, what is left to be explained is the use of *discrete* tone levels in the described L3 French prosody, as French intonation is much more gradient in nature. While the missing piece of the story certainly has to do with Cantonese, it is insightful to discuss it in terms of L2 English with highly relevant L1 Cantonese features, rather than just Cantonese alone. It must be noted, however, that our current data are also compatible with a view that it is only L1 Cantonese which is responsible for the observed non-native French prosody; further research will provide more clues about the extent to which L2 English plays the role of a proxy in language transfer.

L2 English prosody by L1 Hong Kong Cantonese speakers has been analyzed in terms of tonal transfer from Cantonese to English (Luke 2000, Cheung 2008, Gussenhoven 2012, Yiu 2014). Crucially, the patterns of tone assignment depend on word stress in English. The following two examples are from Luke (2000):

English word	Tonal pattern using Cantonese tones
examination	M-H-H-H-L
encyclopedia	M-M-M-H-L

Luke (2000) describes the use of the general tonal template of Mid-High-Low (from Cantonese mid level, high level, and low level tones, respectively). There appears to be an algorithm that could account for tone assignment. Give an English word, the syllable with the primary stress is assigned H (for a high level tone): *examination*, *encyclopedia*. The syllable with the secondary stress may or may not bear H: *examination* with H, *encyclopedia* without H. Then, all syllables sandwiched between two H's are assigned H's: *examination* with H; see Cheung (2008) and Yiu (2014) on high tone spans in Hong Kong English. After all H's are assigned, L (for a low level tone) and M (for a mid level tone) are assigned. All unassigned syllables to the left of the first H bear M's (*examination*, *encyclopedia* with M), and all unassigned syllables to the right of the rightmost H bear L's (*examination*, *encyclopedia* with L).

The use of Cantonese tones in L2 English has important ramifications for our understanding of the L3 French prosody. First, the tonal patterns in L2 English demonstrate the tendency that L1 Cantonese speakers use Cantonese tones in their non-native languages. Second, the precise patterns of which Cantonese tones are used at which syllables or words in the target language depend on prosodic properties of the target language, particularly in terms of pitch. Third, no syllables in the target language are toneless; this conforms to the property of L1 Cantonese that all syllables bear a tone. In the target language, if a syllable is not assigned a tone, it is assigned one by a certain tone-spreading mechanism.

We are now at a position to provide an analysis of L3 French prosody by Hong Kong speakers. The analysis is in terms of language transfer, drawing features from both L1 French and L2 English discussed above. The two major components of the analysis are the assignment of H and L tones, and high tone spanning.

Let us consider the French phrase *la capitale touristique* “the touristic capital” used in Figure 2. First, function words (most of which are monosyllables) are assigned L. As for H, because most (non-sentence-final) phrases in L1 French have a rising pitch contour, and because the phrases usually end with a content word, it is reasonable to assume that all content words in L3 French have their final syllable assigned with H (recall that L2 English assigns H to stressed syllables, but French has no word stress). In other words:

L H H
 | | |
 la capitale touristique

Our view that it is the final syllable of a content word which bears H is paralleled by the analysis of Central African French (Bordal 2015). This variety of French has been in close contact with Sango, a tone language. The tonal pattern for content words in Central African French is (L+)H – with H on the final syllable and L on all preceding syllables in the case of a polysyllabic word, or just H for a monosyllabic word. The case of Central African French provides support to the idea that there is a tendency for non-native French influenced by a tone language to have final syllables of content words bear some sort of a high tone.

Since no syllables can be toneless in our case of L3 French, high tone spreading---similar to L2 English---assigns H to toneless syllables in content words, formally by spreading from the final-syllable H:

L H H
 | / /
 la capitale touristique

The word is the domain of high tone spreading. This results in exactly what is observed as reported in Figure 2 above:

L H H H H H H
 | | | | | | |
 la capitale touristique

Finally, the fact that L3 French assigns tones by the division of function and content words may possibly coincide with a presumably universal tendency for content words being uttered at a higher pitch (and therefore more loudly, due to the close correlation between pitch and loudness in human speech) because of their higher importance for discourse information load, in contrast with function words uttered at lower pitch and more softly. Functional factors may therefore also be at work in shaping L3 French prosody.

CONCLUSION AND IMPLICATIONS

This paper has provided a description of the non-native French prosody by native speakers of Hong Kong Cantonese. Given the socio-historical background of Hong Kong, the speakers are also competent in English as a second language, and French can be considered a third language. The striking prosodic pattern in L3 French by Hong Kong speakers is that the Cantonese high level tone is used for syllables of French content words, and the Cantonese low level tone for syllables of French function words. Such a pattern has been accounted for in terms of interlanguage transfer, from L1 Cantonese to L3 French via L2 English interlanguage.

In terms of language learning and teaching, by studying a target language other than English, this paper demonstrates the general strong tendency for tone language speakers to employ tone in learning non-native languages. Anecdotal comments suggest that the speakers themselves are in general unaware of the specific prosodic patterns in their target language. Therefore, making learners aware appears to be important so that they can more actively attempt to achieve more target language-like prosody. To this end, visualization techniques, discussed by Levis and Pickering (2004) and others, are appropriate for precisely the dual purposes of raising learners' awareness, on the one hand, and providing visual and instant feedback for improvement, on the other. Visualization of prosody has been shown to be helpful for non-tone language speakers learning a tone language (e.g., Chun et al. 2013 with native English speakers learning Mandarin Chinese). For the opposite situation of tone language speakers learning a non-tone language---a case of which is reported in the present paper, it is expected that visualization techniques for prosody will help learners steer away from the use of discrete tone levels and imitate the much more gradient prosodic features in the target non-tone language.

ACKNOWLEDGMENTS

We are grateful for the questions, comments, and encouragement from the audience of the conference held at the University of California, Santa Barbara, on September 5-6, 2014. Guri Bordial Steien and two anonymous reviewers provided insightful comments. We also thank those who have provided us feedback earlier, especially Vincie Ho and K. K. Luke, as well as the undergraduates of the French language program at the University of Hong Kong who kindly agreed to be recorded for the cause of science.

ABOUT THE AUTHORS

Jackson L. Lee (jsllee@uchicago.edu) is a graduate student in linguistics at the University of Chicago. He obtained his BA in Linguistics and French from the University of Hong Kong, and MA in Linguistics from the University of Manchester. He is a computational linguist, focusing

on morphology and phonology. With general interests in language acquisition and the lexicon, his doctoral dissertation focuses on morphological paradigms from the perspective of unsupervised learning. He has long-term interests in tone and intonation. For a better understanding of the lexicon, he has also worked on topics on Cantonese grammar ranging across phonology, morphology, syntax, semantics, and grammaticalization.

Stephen Matthews (matthews@hku.hk) is Professor of Linguistics at the University of Hong Kong and Co-director of the Childhood Bilingualism Research Center at the Chinese University of Hong Kong. His research interests include linguistic typology, language contact, as well as bilingual and multilingual acquisition. He has published on the grammar of Cantonese and other Chinese varieties, on bilingual first language acquisition, and on third language acquisition. He is a co-author of *The Bilingual Child: Early Development and Language Contact* (2007, Cambridge University Press, co-authored with Virginia Yip) which is the 2009 recipient of the Leonard Bloomfield Book Award by the Linguistic Society of America.

REFERENCES

- Boersma, P. & Weenink, D. (2008). Praat: doing phonetics by computer. Version 5.0. (<http://www.praat.org/>)
- Bordal, G. H. (2015). Traces of the lexical tone system of Sango in Central African French. In E. Delais-Roussarie, M. Avanzi, & S. Herment (Eds). *Prosody and Language in Contact* (pp. 29-49). Berlin: Springer.
- Cenoz, J., Hufeisen, B., & Jessner, U., Eds. (2001). *Cross-linguistic influence in third language acquisition*. London: Multilingual Matters Limited.
- Cheung, W. H. Y. (2008). Span of high tones in Hong Kong English. *HKBU Papers in Applied Language Studies*, 12, 19-46.
- Chun, D. M., Jiang, Y., & Ávila, N. (2013). Visualization of tone for learning Mandarin Chinese. In J. Levis & K. LeVelle (Eds). *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference* (pp. 77-89). August 2012. Ames, IA: Iowa State University.
- Crystal, D. (2010). *The Cambridge Encyclopedia of Language*. Cambridge: Cambridge University Press. 3rd edition.
- Di Cristo, A. (1998). Intonation in French. In D. Hirst and A. Di Cristo (Eds). *Intonation Systems: A Survey of Twenty Languages* (pp. 195-218). New York: Cambridge University Press.
- Goldsmith, J. A. (1976). *Autosegmental phonology*. PhD dissertation, MIT.
- Goldsmith, J. A. (1990). *Autosegmental and metrical phonology*. Blackwell.
- Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge: Cambridge University Press.
- Gussenhoven, C. (2012). Tone and intonation in Cantonese English. In *Proceedings of the Third International Symposium on Tonal Aspects of Languages*. Downloaded from <http://gep.ruhosting.nl/carlos/64.pdf>

- Kellerman, E. (1979). Transfer and non-transfer: where we are now. *Studies in Second Language Acquisition*, 2, 37-57.
- Leung, Y.-k. I. (2007). Third language acquisition: why it is interesting to generative linguistics. *Second Language Research*, 23, 95-114.
- Levis, J. M. (1999). Intonation in theory and practice, revisited. *TESOL Quarterly*, 33, 37-63.
- Levis, J. M. & Pickering, L. (2004). Teaching intonation in discourse using speech visualization technology. *System*, 32, 505-524.
- Luke, K. K. (2000). Phonological re-interpretation: the assignment of Cantonese tones to English words. Paper presented at the 9th International Conference on Chinese Linguistics, National University of Singapore, June 2000.
- Maddieson, I. (2013). Tone. In M. Dryer & M. Haspelmath (Eds). *The World atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (<http://wals.info/chapter/13>)
- Matthews, S. & Yip, V. (2011). *Cantonese: A comprehensive grammar*. London: Routledge. 2nd edition.
- Post, B. (2000). *Tonal and phrasal structures in French intonation*. The Hague: Thesus (Subsidiary of Holland Academic Graphics).
- Yip, M. (2002). *Tone*. Cambridge: Cambridge University Press.
- Yip, V. & Matthews, S. (2012). Learner corpora and cross-linguistics influence in third language acquisition. Presented at the International Conference on Bilingualism and Comparative Linguistics, Hong Kong, China, May 15-16, 2012.
- Yiu, S. (2014). Tone spans of Cantonese English. In *Proceedings of the Fourth International Symposium on Tonal Aspects of Languages*.
- Yuen, I. (2007). Declination and tone perception in Cantonese. In T. Riad & C. Gussenhoven (Eds). *Tones and Tunes: Experimental Studies in Word and Sentence Prosody* (pp. 63-78). Berlin: Mouton de Gruyter.

ARE FRENCH IMMERSION “ACCENTS” UNIQUE?

Livia Poljak, Simon Fraser University

The purpose of this investigation was to determine whether or not the British Columbia French immersion program produced differing accents (in this case defined as pronunciation) from other L2 French programs in the province. Five native speakers of French teaching in the Department of French at Simon Fraser University evaluated words, sentences, and narrative utterances of 17 L2 French speakers living in a non-Francophone environment, who completed either high school French immersion or Core French (FSL). Using a program ID choice task, listeners indicated which program the anonymous speakers had completed. Results suggested that French immersion speakers can be distinguished from Core French speakers at above chance levels, though success rates among listeners varied slightly. Formal analysis demonstrated that longer utterances produce more accurate choices. Self-reports of immersion speakers suggest that they spent a greater amount of time with their immersion peers both inside and outside the school environment than with English-speaking peers.

INTRODUCTION

French immersion (FI) is an L2 French program offered in many schools around the world. It originated in Canada in the early 1970s (Roy, 2008) based on the bilingual education theories of McGill University professors Penfield and Lambert (Fraser, 2011). Typically, participants in this elementary and high school program are all L2 learners of French who start off with no knowledge of French (Lambert et al, 1983). The students are meant to acquire the language through interaction rather than learn through formal instruction, and become bilingual by the end of the program (Lapkin et al. 1983). Previous research has focused on lexical, syntactical or morphological mastery among FI students, group dynamics and social interaction among students, and most recently, some VOT analysis (Genesee, 1978; Flege 1995; Courcy 2002; Mougeon et al. 2004; Birdsong 2004, 2007; Netelenbos, 2013). Furthermore, prior research has also suggested that FI students had closer interpersonal relationships with their immersion classmates than with their Anglophone peers (Courcy, 2001). However there is a lack of data on FI accents.

Current Question

This study was part of a larger body of work where I hoped to begin to answer the broad question: does FI lead students to develop their own French “Immersionese” accent? The study defined accent in terms of pronunciation. One objective was to examine whether or not FI speakers sounded different from Core French (CF) to L1 French listeners. I therefore asked the following question:

- Can L1 French listeners distinguish FI from CF speakers based on L2 accent alone?

METHODS

The full study consisted of two perception tasks (an accent rating task, and a program ID choice task) and an acoustic analysis (VOT and formant analysis). Here, attention will be placed on the results of the program ID choice task. In this task, L1 French listeners had to choose which program they believed each speaker had completed, based only on their pronunciation of words, sentences and extemporaneous utterances.

Speakers

Speech samples were collected from a total of 22 speakers who had completed a variety of French high school programs in British Columbia. Among those, 17 (4 male, 13 female) speakers who completed early French immersion (EI), late French immersion (LI), and the FSL program known as Core French (CF). Furthermore, in this study, all students not attending the FI program will be referred to as “English program peers” because the majority of their education is taught in English. Program differences according to the BC Ministry of Education are summarized in Table 1.

Table 1

Key Differences Between French Programs in BC

French Program	Starting Year	% of French Used in Elementary School	% of French Used in High School
EI	K	100-80	75-25
LI	6	100-80	75-25
CF	4	7	7

An online questionnaire was used to elicit background information from speakers. All speakers were between 19 and 23 years of age, and had completed a high school French program in BC in the last two years. A majority of FI speakers were monolingual English speakers and half of CF speakers were L1 English speakers. The online questionnaire also provided information on the interpersonal relationships among FI speakers.

Elicitation

Written French instructions were presented on a computer screen prior to elicitation, and oral instructions were given in English. An interchanging red and green computer screen would signal to the speaker when to listen (red screen) and when to speak (green screen). The recordings took place in a sound-treated room in the Applied Phonetics Lab in the Department of Linguistics at Simon Fraser University. The speaker recordings were digitally edited to separate all word, sentence and extemporaneous stimuli.

Speech Materials

Speakers recorded words and sentences heard during two delayed repetition tasks, as well as an extemporaneous narration using a picture story (Derwing and Munro, 2013). Some of the words had previously been used by Birdsong (2003) in an analysis of L2 French among adult learners of French. All speakers recorded a total of 10 words, each with a particular target sound for listeners to focus on (see Table 2). From these, three words *temps* (/t/), *lundi* (/ɛ̃/) and *bureau* (/y/) were selected for analysis. The English and French /y/ differ in terms of vocal frequency, and nasal vowels such as ɛ̃/ as more common traits of the French language, so English speakers may tend to fully pronounce the /n/. There are also differences in VOT and aspiration between the French and English /t/, and although this sound does not impede intelligibility, it may come into play in terms of accentedness. Only these three sounds were selected to save time for listeners.

Table 2

Words and Target Sounds for the Program ID Choice Task

Words	Target Sounds
bureau	/y/
château	/a/
compliment	/ɑ̃/
coup	/u/
lundi	/ɛ̃/ or /œ̃/
pain	/ɛ̃/
père	/ɛ/
pré	/ʀ/
temps	/t/
tombeau	/o/

Speakers also recorded 7 sentences, of which 3 (“*Il y a un tombeau au milieu du pré.*”, “*Lundi, si j’ai le temps.*”, “*J’ai laissé les documents à mon bureau.*”) were later analysed by listeners. As indicated by the underlined words, sentences used words from the word repetition task. Finally, using a picture story (Derwing and Munro, 2013) with 6 images depicting two individuals accidentally exchanging suitcases, speakers each gave their own versions as to what happened in the story.

Listeners

The 5 listeners were all L1 French speakers living in BC, but were originally from France or Quebec. They were post secondary professors in the French Department at Simon Fraser University, between the ages of 36 and 60. All listeners had also lived in other English speaking countries or provinces in Canada prior to arrival in BC. Listeners also claimed to be familiar with the French programs offered in BC and had between 14 and 35 years of French teaching experience. When asked about the importance of pronunciation for French proficiency, listeners gave mixed responses ranging from 7 (very important) to 5 (somewhat important). When queried about their impressions of FI students' accents, 4 of the 6 replied that they thought FI students have similar accents to each other, but not to other French program students, and two listeners expected that FI students would sound similar to Francophone students from the Programme Cadre school system.

Listening Procedure

Using a Praat script, listeners evaluated the stimuli in the same sound-treated room. Listeners attempted to guess the program each speaker had completed, first based on their pronunciation of words. For the program ID choice task, listeners first heard and evaluated all speakers saying *temps* in a randomized order, followed by *lundi* and finally *bureau*. Then, listeners repeated the choice task then for three sentence sets, S1, S4, S7, (randomized) and finally for the narration (randomized). There was a possibility for the listeners to replay each word and sentence up to 3 times before they made their evaluation, after which the next stimulus was presented. Because previous formal analysis (not presented here) on CF and FI accent ratings had demonstrated no significant difference between EI and LI speakers, the two groups were combined. Therefore listeners only had to choose between FI and CF groups for the choice task.

RESULTS

Web Survey Results

While there is some anecdotal evidence that listeners can distinguish between FI and CF speakers, the question remains as to why. As was suggested in a case study on FI students (Courcy 2001), they tended to spend more time with other FI peers in school even during break time, than with English speaking peers, and going so far as to use French as a means to distinguish themselves from their English-speaking peers. As this could be a potential explanation for a common accent among FI students, participants in this research were asked how much time they spent with their FI peers in comparison with their English program peers in a) the school setting; and b) outside of school. Only EI and LI students had to answer these two questions.

It was again found that typically FI speakers spent much more time with their FI peers in school than with English program peers (see Table 3), with LI speakers claiming to spend the most time with their FI peers. These findings are similar to Courcy (2001), showing that, at least in the school environment, FI students were closer to their FI peers than to students outside the program.

Outside the school setting, LI speakers still spent the most time with their French program peers (see Table 4). In fact, 6 out of 7 EI speakers claimed to spend at least an equal amount of time with their FI peers as with their English program peers. These findings seem to corroborate Courcy's case study results.

Table 3

Amount of Time Spent with Immersion Peers vs. English Program Peers in School (Total Number of Immersion Speaker Self Ratings)

French Program	Much More Time	A Little More Time	Equal Amount of Time	A Little Less Time	A Lot Less Time	Total
Early	5	0	1	1	0	7
Late	4	0	0	0	0	4

Table 4

Amount of Time Spent with Immersion Peers vs. English Program Peers Outside of School (Total Number of Immersion Speaker Self Ratings)

French Program	Much More Time	A Little More Time	Equal Amount of Time	A Little Less Time	A Lot Less Time	Total
Early	2	3	1	0	1	7
Late	3	1	0	0	0	4

Program ID Choice Task Results

Listeners were able to correctly identify the speakers' programs from the recorded words a majority of the time. The figures show the accuracy of each listener by category and demonstrate that in all cases except one, the listeners correctly identified speaker categories with over 50% accuracy for all three word tokens. Furthermore, listeners had an accuracy rating of 67%, 68% and 65% respectively for each of the token words seen below.

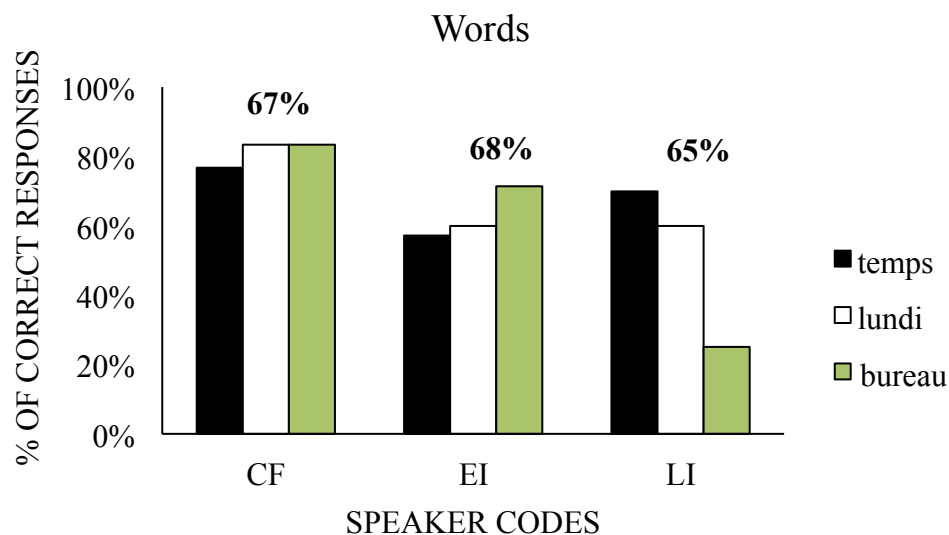


Figure 1. Correct speaker program choices by word for all speaker groups. Percentages indicate the overall accuracy for all three speaker groups.

Statistical analysis results with $p=0.05$ or below, indicate significantly better than chance performance (see Table 5). A binomial distribution was used to evaluate whether each total was statistically better than chance. Listeners were able to identify speakers' programs 40% of the time (6/15) at above-chance levels on the basis of isolated word production.

Table 5

Binomial Probability Results for Token Words, Sentences and Extemporaneous Narration

Item	J871	J729	J987	J121	J678
<i>temps</i>	0.1855	0.0944	0.0944	0.0472*	0.0052**
<i>lundi</i>	0.1484	0.0052**	0.0944	0.0472*	0.0944
<i>bureau</i>	0.0944	0.1484	0.0472*	0.1855	0.0182*
S1	0.0944	0.0052**	0.0010***	0.0182*	0.0182*
S7	0.1484	0.0010***	0.0052**	0.0472*	0.0944
S4	0.0472*	0.0001***	0.0052**	0.0182*	0.0010***
Narration	0.0472*	0.0052**	0.0182*	0.0052**	0.0010***

* $p < 0.05$

** $p < 0.01$

*** $p < 0.005$

Longer stimuli demonstrated more accurate results from listeners for the program ID choice task. This can be seen in the program ID choice task results for the sentence productions. Listeners were able to significantly identify speaker programs 80% of the time for sentences. This was especially true for CF speakers (90% accuracy for S1, 100% for S7 and 97% for S4). In total, listeners were accurate 76% of the time for S1, 80% of the time for S4, and 70% of the time for S7 (see Figure 2). Listeners were accurate above chance 70% of the time for sentences (see Table 5).

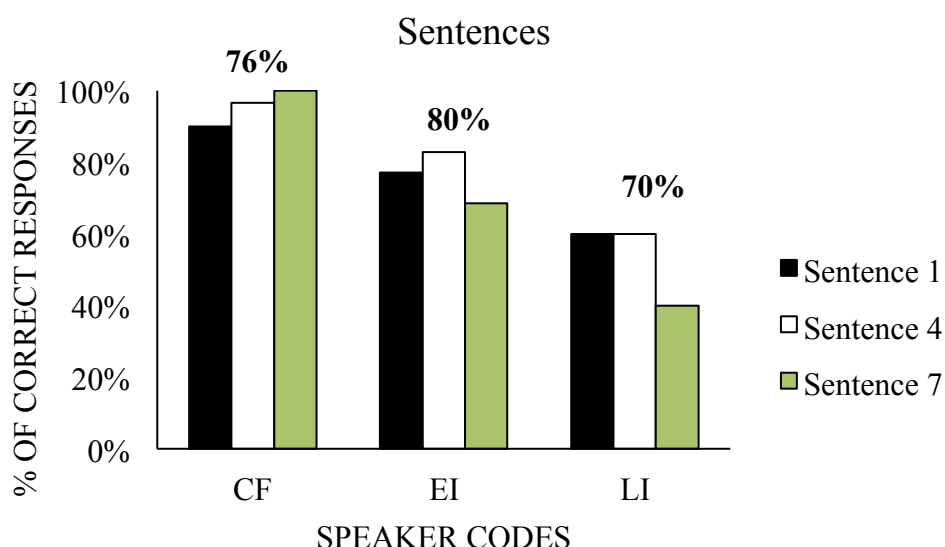


Figure 2. Correct speaker program choices by listeners for *S1*. Percentages indicate the overall accuracy for all three speaker groups.

Finally, the extemporaneous results (Figure 7) were very comparable to the above sentence results. Listeners were best at guessing what program CF speakers completed (93% accuracy) followed by EI (77%) and the LI (65%). In total listeners were above chance 100% of the time for the narration task (see Table 5). Both sentence accuracy results and narration accuracy results seem to confirm that it was easier for the listeners to identify programs when utterances were longer.

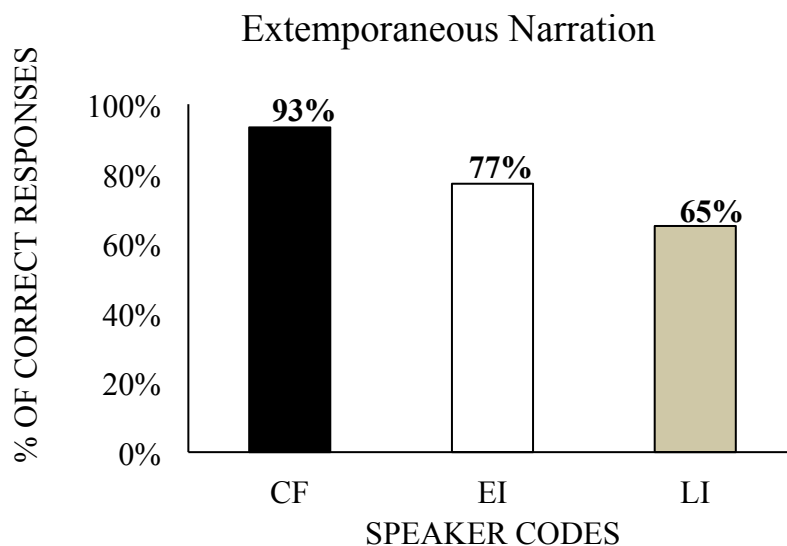


Figure 7. Correct speaker program choices by listeners for Extemporaneous Narration. Percentages indicate the overall accuracy for all three speaker groups per listener.

Some differences in success between listeners were noted for this task. Listener J871, for instance, was only able to differentiate significantly between speakers for *S4* and the narration task (see Table 5). On the other hand, listener J121 had the highest success at differentiating between speaker groups. Further analysis suggests that some speakers were more readily identified as FI or CF than others (see Table 6). Interestingly, CF speakers were more frequently evaluated as being CF than FI speakers were evaluated as FI. Only two FI speakers (EI1469, LI1847, marked with an asterisk in Table 6) were labeled more readily as CF. Of these two, LI1847 was most often mistaken for a CF speaker.

Table 6

Percentage of Times Each Speaker's Program was Correctly Identified

Speakers	words	sentences	narration
CF1806	93%	93%	100%
CF5844	60%	93%	100%
CF7209	93%	93%	60%
CF7716	93%	100%	100%
CF8024	47%	93%	100%
CF9685	100%	100%	100%
EI1469*	73%	40%	40%

EI2080	53%	87%	80%
EI2523	27%	80%	60%
EI2893	93%	87%	100%
EI2981	73%	80%	80%
EI8048	53%	87%	80%
EI8550	67%	73%	100%
LI1847*	33%	27%	20%
LI2009	47%	80%	100%
LI5481	40%	53%	60%
LI9778	87%	53%	80%

DISCUSSION

Results indicate that listeners' program guesses were more accurate when utterances were longer than a single word, though even guesses based on isolated words were sometimes accurate. It should also be noted that total results were above 50% accuracy for words, sentences and narrations. In general, EI and CF program ID choice results showed that listeners could identify with moderate accuracy what program these speakers finished. One possible reason for this is that FI students have more experience with French, and thus had more time to improve their accents, leading listeners to simply choose FI for those speakers whose accents sounded more native-like, and CF for speakers who sounded less native-like.

Additionally, all listeners were able to distinguish between FI and CF at above-chance levels for the narration task. This could suggest that listeners were not listening for pronunciation alone, but for a combination of phonetic, prosodic, grammatical and lexical information, as was previously suggested in other works (Genesee, 1978; Tatto, 1983). Together, these traits may have helped listeners identify the program these individuals finished. This was also suggested by some of the listeners upon completion of the program ID choice task.

However, a second possible explanation for why FI students sound different from CF students is that the FI program promotes a sense of community, or "apartness" from other non-FI students. Unlike Francophone program schools, which are often in separate buildings from English program schools, FI programs are normally housed together with the English program. Despite this, it was demonstrated that FI speakers spent more time with their FI peers than with English program peers, even after entering high school where only a maximum of 3 classes are taught exclusively in French. In the rest of the classes, FI students would be mixed with English program students. The results from the online questionnaire concur with the case study by Courcy (2001) where FI students claimed to have closer relationships with their Immersion classmates than with English program peers. A unique FI pronunciation could therefore be a result of group affiliation, where speakers choose to emulate a shared pronunciation pattern, resulting in a homogenised accent. Orr (2011) found anecdotal evidence that homogenised

accents could occur when a common L2 was being used as a tool for communication among differing L1 speakers.

It is possible therefore that the FI program is producing a sense of group affiliation that extends to pronunciation as well. Gatbonton et al (2005) demonstrated that in cases where the minority group is large enough, some speakers chose to sound foreign to prevent appearing “less loyal” to their L1 community. Pronunciation patterns can therefore be used as a marker of group affiliation. FI students may too be choosing to sound more similar to each other than to CF, who may be viewed as outsiders to the “FI community”. To add to this, one of the listeners presented a personal anecdote during their rating task, in which she reported that her son (a Francophone) had been faking an L2 French accent to avoid sounding different from his FI peers. It is unclear if FI students feel isolated from their English program peers because of their French instruction, or if their view their self-earned bilingual status as superior to both monolingual English program students and Francophones and therefore choose to associate less with them in elementary and high school.

However, it should be noted that sounding more native-like does not appear to be a goal, or a necessity to FI speakers in this study. Nor does their pronunciation appear to impede intelligibility. This is perhaps unsurprising, because the FI program is tailored to non-Francophone living in a non-Francophone environment. They do not have to fit in to the French-speaking community, because they may already feel like they are part of the aforementioned “FI community”. What should be highlighted here is that this newly-formed environment is not based on ethnicity or social status, but rather on the program itself. This means that the pronunciation is institutional in nature, and may therefore not exist if the program was not created in the first place. Taking a closer look at FI accents may be gateway towards identifying a new subgroup of accents: Institutional accents.

Altogether, the results of the Program ID Choice task (forced choice FI vs. CF) indicated that the native-French listeners could distinguish FI speakers from CF speakers. In general, longer utterances provided the most accurate results. Although the research analysed only a portion of the recorded words and sentences, the results are an important step toward identifying FI accents as distinct from other L2 French accents. However, further research would have to be conducted on immersion (such as comparing FI accents to Francophone program accents) in order to better conclude if the FI program produces a distinct accent, and thereby also help indicate if an institution alone can foster a distinct L2 accent. Though this study marks the beginnings of the answer, as of yet, this question remains unsolved. I hope this study will further the interest in immersion accents, and immersion group affiliation in the future.

ACKNOWLEDGMENTS

I would like to thank my mentor Dr. Murray Munro (Simon Fraser University, Department of Linguistics) for his enormous help and inspiration. This work would not be what it is without him. I would also like to thank the SFU French Department, for helping me find speakers and listeners. A special thanks also to Camille Panier and Alida Soucé who recorded the tasks that speakers repeated for this research, and to Eric Hedekar for his help with Praat.

ABOUT THE AUTHOR

Livia Poljak (lpa10@sfu.ca) is an MA graduate from the Department of French at Simon Fraser University with a specific interest in L2 French accents. A graduate of late French immersion, she earned an undergraduate degree in political science and French language, and she learned and maintained five languages (English, French, Hungarian, Serbian and Spanish). She will be continuing her studies in examining the accents of French immersion students across Canada. Her interests lie in bilingual L1, L2 and L3 acquisition, language and group affiliation, and L1/L2 accent transfer.

REFERENCES

- Birdsong, D. (2003). Authenticité de prononciation en français L2 chez des apprenants tardifs anglophones : analyses segmentales et globales. *Acquisition et Interaction en Langue Étrangère*, 18, 31 October 2013.
- Birdsong, D. (2007). Native like pronunciation among late learners of French as a second language. In O-S Bohn, & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 99-116). Amsterdam: John Benjamins Publishing Company.
- Courcy, M. (2001). *Learners experiences of immersion education: Case studies of French and Chinese*. Clevedon, GBR: Multilingual Matters.
- Derwing, T. M., & Munro, M. J. (2013). The development of L2 oral language skills in two L1 groups: A 7-year study. *Language Learning*, 63, 163–185.
- Flege, J. E. (1987). The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics*, 15, 47-65.
- Fraser, G. (2011). Immersion schools: Wallace Lambert's legacy. *Canadian Issues*, 5-9.
- Gatbonton, E., Trofimovich, P., Taylor, D. (2005). Learners' ethnic group affiliation and L2 pronuncaiont accuracy: A sociolinguistic investigation. *TESOL Quarterly*, 39, 489-511
- Genesee, F. (1978). A longitudinal evaluation of an early immersion school program. *Canadian Journal of Education / Revue Canadienne de l'Éducation*, 3(4), 31-50.
doi:10.2307/1494684
- Lapkin, S., Swain, M., & Argue, V. (1983). *French immersion: The trial balloon that flew*. Toronto: OISE Press.
- Mougeon, R., Rehner, K., & Nadasdi, T. (2004). The learning of spoken French variation by immersion students from Toronto, Canada. *Journal of Sociolinguistics*, 8(3), 408-438.

- Netelenbos, N. (2013). *The development of voice onset time in French immersion children*. (Unpublished Masters). University of Lethbridge, Lethbridge, Alberta.
- Orr, R., Quené, H., Diefenback, T., van Leeuwen, D. A., Huijbergts, M. (2011), An international English speech corpus for longitudinal study of accent development *Interspeech*, ISCA, pp. 1889-1892.
- Rehner, K., & Mougeon, R. (1999). Variation in the spoken French of immersion students: To ne or not to ne, that is the sociolinguistic question. *Canadian Modern Language Review*, 56(1), 124.
- Tatto Mabel, A. (1983). *A comparative analysis of grammatical errors in the written code between grade eleven immersion French and grade eleven core French*. (Unpublished Master of Arts Teaching of French). Simon Fraser University, Burnaby.

Simonchyk, A., & Darcy, I. (2015). Acquisition of word-final devoicing by American learners of Russian. In J. Levis, R. Mohammed, M. Qian & Z. Zhou (Eds). *Proceedings of the 6th Pronunciation in Second Language Learning and Teaching Conference* (ISSN 2380-9566), Santa Barbara, CA (pp. 85-94). Ames, IA: Iowa State University.

ACQUISITION OF WORD-FINAL DEVOICING BY AMERICAN LEARNERS OF RUSSIAN

[Ala Simonchyk](#), Indiana University

[Isabelle Darcy](#), Indiana University

This study investigated the acquisition of the phonological rule of word-final devoicing by American learners of Russian and examined the effects of articulatory features, such as place of articulation, manner of articulation and palatalization, on the degree of voicing preserved in final obstruents. Twenty-six American learners of Russian participated in a word-learning task to memorize 24 target words and subsequently performed a picture-naming task. In order to control for previous lexical knowledge, frequency effects and phonetic environment, we used pseudowords that were matched to pictures of real objects, which were assigned a new meaning related to space travel. Minimal pairs were excluded to avoid task effects. The results suggested that learners did not fully acquire the rule of word-final devoicing in Russian despite the fact that voiceless consonants are unmarked and occur in English word-finally. Manner of articulation had a significant effect on the degree of voicing. Stops retained more voicing than fricatives.

INTRODUCTION

The phonological rule of word-final devoicing states that voiced obstruents (i.e., stops and fricatives) become voiceless in word-final position. Learners whose native language preserves voicing contrasts word-finally, e.g. English, must learn to discard the feature [+voice] in that position for a neutralizing language such as Russian. The Markedness Differential Hypothesis, proposed by Eckman (1977), suggests that areas of a target language that are different from the native language and more marked cross-linguistically are difficult to acquire. Voiced consonants are also considered more marked than voiceless consonants. Word-final voicing contrasts are more marked than voicing contrasts found in the word-initial or word-medial positions. Since voiceless consonants are not considered marked cross-linguistically, English learners of Russian should not experience much difficulty acquiring the rule of word-final devoicing, especially taking into account the fact that voiceless consonants occur in word-final position in English. There are a few studies that have investigated the acquisition of marked voiced consonants by speakers of neutralizing languages (Cebrian, 2000; Simon, 2010; Smith, Hayes-Harb, Bruss & Harker, 2009), but almost none that have examined the acquisition of unmarked voiceless consonants by speakers of non-neutralizing languages (Dmitrieva, Jongman & Sereno, 2010). Counter to the predictions made by the Markedness Differential Hypothesis, anecdotal evidence from pronunciation classes in L2 Russian suggests that the phonological rule of word-final devoicing is not readily acquired across the whole proficiency spectrum.

Within the framework of Optimality Theory (Prince & Smolensky, 1993), word-final devoicing is analyzed as the effect of the markedness constraint *VOICED-CODA against voiced

consonants word-finally. The markedness constraint, which governs the cross-linguistic tendency towards final devoicing, is more highly ranked than the faithfulness constraint IDENT-IO(voice), which ensures that outputs match inputs in terms of voicing. Consider evaluation of two candidate outputs for the input /zub/ ‘tooth’ in Russian: (a) [zub] violates *VOICED-CODA, but satisfies IDENT-IO(voice); (b) [zup] satisfies *VOICED-CODA, but violates IDENT-IO(voice). In English, as opposed to Russian, IDENT-IO(voice) dominates *VOICED-CODA, which results in the reverse ranking of faithfulness and markedness constraints. When English learners transfer this ranking into Russian, they preserve the phonological contrast between voiced and voiceless consonants word-finally, which is reinforced by differences in acoustic cues such as duration of the preceding vowel, consonant, voicing into consonant and release. For English learners to acquire the rule of word-final devoicing in Russian, they have to re-rank constraints in such a way that the markedness constraint dominates the faithfulness constraint. The goal of our study is to provide empirical evidence by examining the acquisition of word-final devoicing by American learners of Russian.

Literature Review

Dmitrieva et al. (2010) examined the production of word-final obstruents in Russian by three groups of speakers: monolingual Russian speakers, Russian speakers with knowledge of English and American learners of Russian. In their study, participants were asked to read a list of 34 Russian minimal pairs alternating in word-final voiced and voiceless obstruents. The acoustic analysis was based on four durational measurements: preceding vowel, closure / frication, voicing into closure / frication and release. The findings suggested that American learners of Russian and Russian native speakers used different acoustic cues to encode [+voice]. Russian monolinguals devoiced voiced obstruents word-finally but they did not neutralize the obstruents completely: they produced significant differences in their durations of closure / frication and release. However, monolingual speakers did not produce any differences in the durations of voicing into closure / frication, although it is a very important cue for encoding [+voice] since it indicates how long the vocal cords vibrate to create voicing. Monolingual Russian speakers did not use durations of the preceding vowels either. In English, however, vowel duration is the primary acoustic cue to distinguish between voiced and voiceless consonants (Mack, 1982). Unlike monolinguals, Russian speakers with knowledge of English maintained significant differences in the durations of the preceding vowel and voicing into closure / frication, as well as durations of closure / frication and release. Dmitrieva et al. considered this to be the effect of L2 English on L1 Russian. American learners of Russian produced even greater differences in the durations of the preceding vowel, closure / frication, voicing and release than Russian speakers with knowledge of English did, which suggested that learners did not devoice voiced obstruents similarly to Russian native speakers. However, the most proficient American learners of Russian decreased the durational differences between voiced and voiceless consonants and devoiced word-final obstruents more than the monolingual Russian speakers did.

Target words in Dmitrieva et al.’s study were distributed equally with respect to place and manner of articulation, but the results did not provide any insights into whether neutralization depended on articulatory features. Ohala (1983) noted that aerodynamic requirements and airstream mechanisms involved in the production of obstruents determine how the degree of voicing can be affected by articulatory features. Devoiced fricatives are more common than devoiced stops because voiced fricatives have more exacting aerodynamic requirements and

require greater glottal airflow to maintain voicing than voiced stops do. In order to sustain voicing, oral pressure should be low, whereas in order to sustain frication, oral pressure should be high. As for the place of articulation, Ohala (1983) mentioned that labial stops have greater compatibility with voicing than velar stops, since the oral cavity is much smaller for the latter than for the former. Constriction for /b/ is at the lips, whereas constriction for /g/ occurs at the velum. Vibrations of the vocal cords are the source of voicing. Therefore, the distance from the vocal cords to the lips is greater than to the velum.

Following the same line of reasoning, we would also expect that palatalized consonants will be more devoiced than their plain counterparts on the assumption that the production of palatalized consonants requires the tongue to press against the palate, which creates a smaller oral cavity. When producing a plain /b/, the tongue is low and flat and the oral cavity is wide. However, for a palatalized /bʲ/, the body of the tongue is raised, and as a result the oral cavity becomes smaller and the channel is narrower. Therefore, a plain /b/ is expected to have more voicing than a palatalized /bʲ/. There is no study to date that examines the interaction between palatalized consonants and devoicing. Our study addresses this gap in the literature with an examination of the acquisition of word-final devoicing in Russian by American native speakers, thereby establishing what effect articulatory features can have on the feature [+voice].

Research Questions

The following questions guided the current investigation:

1. Do low-intermediate American learners of Russian devoice word-final voiced obstruents in Russian?
2. Do articulatory features, such as place of articulation, manner of articulation and palatalization, have an effect on the degree of voicing word-finally?

We hypothesize that low-intermediate American learners of Russian will transfer word-final voiced obstruents from English into Russian for a number of reasons. First, according to the Full Transfer / Full Access Hypothesis (Schwartz & Sprouse, 1996), the initial state of interlanguage is the grammar of the native language. Since our participants are low-intermediate, we would expect the effect of their native language to be relatively strong. Second, Russian orthography spells out underlyingly voiced obstruents, similarly to English. This might mislead learners into believing that Russian allows voiced obstruents word-finally. Third, the degree of voicing in word-final obstruents can vary depending on the phonetic environment, functional load and pragmatic reasons (Port & Crawford, 1989). Consequently, American learners may receive input from Russian native speakers, e.g. teachers, with varying degrees of neutralization, including word-final voiced or partially-voiced obstruents in Russian.

With respect to the second research question, it is hypothesized that place of articulation, manner of articulation and palatalization will have an effect on the production of word-final voiced obstruents. We expect that: (a) fricatives will be more devoiced than stops; (b) labials and coronals will retain more voicing than dorsals; and (c) palatalized consonants will be more readily devoiced than their plain counterparts.

METHOD

Participants

The participants of the study were 26 native speakers of American English, adult learners of Russian (11 females, 15 males). The participants' ages ranged from 18 to 35 years old, with a mean age of 23. The age of initial Russian instruction ranged from 15 to 28 years old. The participants were enrolled in levels two and three of an intensive Russian summer program that offers instruction at nine levels. Enrollment in levels was based on the results of an in-house placement test and previous experience with the language. Overall, the level of participants can be characterized as low-intermediate.

Materials

We created 20 target words and 4 distractors. All items had a CVC structure. The onsets were either stops or fricatives. The nuclei were either /u/ or /o/. The codas alternated in target stops /p, pʲ, b, bʲ, t, tʲ, d, dʲ, k, g/ or fricatives /f, fʲ, v, vʲ, s, sʲ, z, zʲ, ʃ, ʒ/. The dorsals /k, g, ʃ, ʒ/ do not have palatalized counterparts in Russian.

Pseudowords were used because the Russian lexicon did not have enough real words of a CVC structure with the word-final target consonants. The target items were matched to pictures of real objects that were assigned meanings related to the topic of space travel (Figure 1). The target words and distractors were divided into two lists to facilitate the task for participants, who had to memorize 12 words instead of 24. This also helped avoid minimal pairs alternating in plain and palatalized consonants. Thus, each list had four labials, four coronals, two dorsals and two distractors. Six words in each list contained a word-final palatalized consonant, e.g., list 1 had /zop/ and /dosʲ/ whereas list 2 had /zopʲ/ and /dos/. In order to avoid task effects, minimal pairs, such as /zop/ - /zob/, were avoided; instead, /zop/ alternated with /kob/. Participants were randomly assigned to either of the lists with an equal number of participants allotted to each list. Pictures and memorization were used to reduce the possible effect of orthography during the elicitation stage (Kharlamov, 2014; Port & Crawford 1989).



Figure 1. Examples of target words with the pictures and their assigned meaning.

Procedure

The experiment took place in a language lab and lasted about 30 minutes. At the beginning of the experiment, the participants were told that they were going on a space trip and had to learn the names of objects that they would need in their space travel.

The word-learning stage A included four exercises. The DMDX software (Forster & Forster, 2003) was used to present auditory and visual stimuli and record participants' responses. In exercise 1 the participants saw pictures with Russian words and their meanings written in English. They also heard the target words produced by a female Russian native speaker. Each picture was presented for 15 seconds (Figure 1). The participants were asked to memorize the words and their associated pictures. In exercise 2 the participants saw a picture and two Russian words at the bottom. The task was to choose the word that described the picture by pressing "1" or "2". In Exercise 3 the participants saw a picture and a question: "Is it (target word)?" at the top. The task was to decide whether the word in the question matched the picture by pressing "no" or "yes". The participants had two seconds to make their choice. Immediate feedback was provided in both exercises, and the target word was repeated for incorrect answers. In exercise 4 the participants saw a picture and were asked to say a matching word out loud within two seconds. Then the participants saw a picture with a word again and heard its pronunciation.

In stage B the participants performed a picture-naming task. The participants saw a picture and were asked to produce a word that matched the picture using a carrier phrase "Это ф... Это ф..." (e.g., This is f... This is f...). The first letter of the word was provided to facilitate retrieval. If the participants could not recall a word, they were shown the same picture with a sentence, e.g., "Я читаю. Это фыр. Это фыр." (I am reading. This is foog. This is foog). The participants were asked to read the sentence. Their productions were recorded using the software Praat (Boersma & Weenink, 2011). After the recording the participants were asked to fill in a language background questionnaire.

Data Analysis

The productions of the target words were coded based on whether they were produced from memory or by reading a sentence in order to track the potential effects of reading on participants' productions. Four durational measurements were collected from each token: (1) preceding vowel; (2) consonant (closure or frication); (3) voicing into consonant; (4) release (only for stops). The boundaries of each durational measurement were established using textgrids in Praat. Then a Praat script was run on the textgrids to extract the measurements. Voicing ratios were calculated for word-final obstruents using the following formula.

$$\text{Voicing ratio} = \frac{\text{Duration of voicing into consonant}}{\text{Duration of consonant}} \times 100\%$$

Voicing ratios showed how much voicing each consonant retained. Theoretically, a fully voiced consonant had a voicing ratio of 100%, whereas a fully voiceless consonant had a ratio of 0%.

The analysis revealed that the participants were not able to produce all the target words from memory; they read 185 tokens (36%) out of 510. A series of one-way ANOVAs was run in SPSS on the durations of the preceding vowels, consonants, voicing into consonants and release produced in voiced consonants with the within-participants factor of mode of elicitation (memory or reading). The results showed that there was no significant effect of mode of elicitation on the durational measurements of voiced consonants. Therefore, it was decided to

group all tokens for analysis irrespective of the mode of elicitation because an effect of orthographic exposure was not found in the experiment.

RESULTS

The first research question asked whether low-intermediate American learners of Russian devoiced word-final voiced obstruents. The durational measurements obtained from learners' productions demonstrated expected tendencies in the manipulation of acoustic cues to differentiate underlyingly voiced and voiceless consonants, suggesting that they did not devoice voiced obstruents in word-final position. Voiced consonants had longer preceding vowels and durations of voicing into consonants, whereas voiceless consonants had longer durations of consonants and release. A series of one-way ANOVAs was run on the durational measurements to establish whether American learners distinguished voiced and voiceless obstruents in their production (Table 1).

Table 1

Mean durational measurements and standard deviations (in parentheses) for underlyingly voiced and voiceless obstruents.

Duration	Voiced (ms)	Voiceless (ms)	<i>p</i> values
Preceding vowel	181 (56)	161 (38)	.001*
Consonant	176 (112)	218 (114)	.004*
Voicing into consonant	37 (28)	8 (12)	.000*
Release	109 (61)	119 (69)	.434

Note. * $p < .05$

Low-intermediate American learners of Russian produced significantly longer preceding vowels $F(1,255) = 11.71, p < .001$, shorter consonants $F(1,255) = 8.57, p < .004$, and more voicing into consonants $F(1,255) = 114.16, p < .000$ for underlyingly voiced obstruents. The learners did not employ release durations as an acoustic cue to distinguish voiced and voiceless consonants. Given that learners distinguished voiced and voiceless consonants for three out of four acoustic dimensions, we can conclude that American learners did not successfully devoice word-final voiced obstruents in Russian.

The second research question asked whether articulatory features had an effect on the degree of retained voicing in underlying voiced consonants. A series of one-way ANOVAs was run on the durations of the preceding vowels and voicing ratios only for voiced consonants with the within-participants factors of manner of articulation, place of articulation, and palatalization. The results showed that vowels were significantly shorter before voiced stops (163 ms) than before voiced fricatives (199 ms), $F(1,125) = 14.16, p < .000$, whereas voicing ratios were significantly longer in voiced stops (61%) than in voiced fricatives (10%), $F(1,125) = 144.93, p < .000$ (Table 2).

Table 2

Mean vowel durations and voicing ratios with standard deviations (in parentheses) for voiced stops and fricatives.

Duration	Stops		Fricatives	
	[+voice]	[-voice]	[+voice]	[-voice]
Preceding vowel (ms)	163 (44)	147 (34)	199 (61)	175 (37)
Voicing ratio (%)	61 (33)	9 (16)	10 (8)	3 (5)

The voicing ratio of voiced fricatives was very low in comparison to that of voiced stops and close to the ratio for voiceless fricatives, which suggested that the participants might have devoiced fricatives more successfully. A two-way ANOVA was run on the durations of the preceding vowels and voicing ratios with the within-participants factors of voice and manner of articulation to identify whether there was a significant difference between voiced and voiceless fricatives and stops. There was a main effect of voice, $F(1,253) = 13.3$, $p < .000$, and of manner, $F(1,253) = 32.41$, $p < .000$, on vowel durations, but there was no interaction between voice and manner. Indeed, vowels were longer before voiced obstruents. They were also longer before fricatives than before stops. There was also a main effect of voice $F(1,253) = 161.15$, $p < .000$ and of manner, $F(1,253) = 146.29$, $p < .000$, on voicing ratios, which means that voicing ratios were larger in voiced obstruents (as opposed to voiceless ones), and in stops (as opposed to fricatives). There was also a significant interaction between voice and manner $F(1,253) = 95.51$, $p < .000$. Voiced stops displayed a higher voicing ratio than voiced fricatives, $F(1,253) = 236.33$, $p < .000$, but voiceless stops vs. fricatives did not differ. Similarly, for fricatives and stops, the feature [+voice] led to higher voicing ratios, $F(1,253) = 4.25$, $p < .04$ and $F(1,253) = 253.4$, $p < .000$, respectively. Therefore, although it seems that the learners devoiced underlying voiced fricatives, in reality they differentiated both voiced and voiceless fricatives using vowel durations and voicing ratios.

Place of articulation had an effect only on vowel durations $F(2,124) = 5.19$, $p < .007$. Post hoc analysis using Tukey's HSD indicated that vowels before labials were significantly shorter than before coronals ($p = .005$). There was no significant difference in vowel durations between coronals and dorsals. Voicing ratios were not significantly affected by place of articulation (Table 3).

Table 3

Mean vowel durations and voicing ratios with standard deviations (in parentheses) for voiced labials, coronals and dorsals.

Duration	Labial	Coronal	Dorsal	<i>p value</i>
Preceding vowel (ms)	164 (50)	198 (58)	180 (53)	.007*
Voicing ratio (%)	38 (38)	36 (33)	31 (33)	.741

Note. * $p < .05$

No statistically significant differences were found in the durations of the preceding vowels or voicing ratios in plain and palatalized consonants. One of the reasons can be that palatalization is a challenging articulatory feature and low-intermediate learners were not able to produce it.

DISCUSSION

The study set out to examine the acquisition of the word-final devoicing rule by low-intermediate American learners of Russian. Although voiceless consonants are unmarked and American learners have the feature [-voice] in their native language, the results suggested that moderately proficient learners of Russian did not devoice word-final voiced obstruents. Underlyingly voiced consonants displayed significantly longer vowel durations and voicing into consonants, as well as significantly shorter consonant durations, which are characteristics of voiced consonants. Differences in release durations were not statistically significant.

Learners might have failed to devoice voiced obstruents for a number of reasons. First, they might have been influenced by their native English phonology, which allows voicing contrasts word-finally. In Optimality Theory terms, learners favored the faithfulness constraint over the markedness constraint. Also, Russian orthography which spells out underlyingly voiced obstruents might have misled learners into believing that Russian has word-final voicing contrasts. Learners could have been aware of the word-final devoicing rule, since they have been learning Russian for more than a year on average. However, they have not fully incorporated it in their second language phonology, at least in production, which suggests that unmarked categories can be as hard to acquire as marked.

The study also attempted to establish the effect of articulatory features, such as manner of articulation, place of articulation and palatalization, on the degree of voicing preserved in voiced consonants word-finally. Manner of articulation seems to have a strong effect on the degree of voicing in word-final consonants. Voiced fricatives have much smaller voicing ratios than stops do, which supports Ohala's claim that "to the extent that the segment retains voicing it may be less of a fricative, and if it is a good fricative it runs the risk of being devoiced" (Ohala, 1993, p. 201). However, despite low voicing ratios, there were significant differences between underlyingly voiced and voiceless fricatives, which indicated that learners do not actually neutralize fricatives. Variation in the degree of voicing in word-final fricatives observed in the productions of American learners can be explained by the airstream mechanism and aerodynamic characteristics of fricatives, rather than by learners' application of the word-final devoicing rule. Moreover, learners used durations of the preceding vowel as an additional way to encode [+voice]. In conclusion, the ability to devoice word-final voiced consonants, even fricatives, which are more susceptible to devoicing, does not seem to be easily acquired by native speakers of non-neutralizing languages such as English.

The finding that manner of articulation has a strong effect on the degree of preserved voicing carries an important methodological implication for studies that seek to investigate voicing neutralization. Data from test materials that include stops and fricatives should be analyzed separately. Otherwise, the results can be skewed in favor of devoicing if fricatives are used in most of the test tokens or, vice versa, in favor of voicing, if stops represent the majority of the tokens.

Place of articulation had a significant effect on preceding vowel durations but not voicing ratios.

However, raw data show that labials have larger voicing ratios than dorsals, which supports Ohala's claim that labials can sustain voicing better than dorsals. Palatalization did not have an effect on the degree of voicing. However, recruiting advanced participants in a future study might reveal this effect more clearly because advanced participants are likely to be more accurate at producing palatalized consonants than low-intermediate learners.

Future research can be directed in two ways. Recruiting advanced participants will answer the question as to whether American learners of Russian learn to devoice word-final voiced obstruents at later stages of acquisition. It will also test the effects of palatalization more efficiently. The second direction will involve recruiting native speakers of Russian and English to investigate the effects of articulatory features on the feature [+voice] in neutralizing and non-neutralizing languages. It will help establish whether such effects are language-general, language-specific or only observed in interlanguage.

ABOUT THE AUTHORS

Ala Simonchyk (alisimon@uemail.iu.edu) is a PhD student in the Department of Slavic and East European Languages and Cultures and the Department of Second Language Studies at Indiana University. She teaches Russian Pronunciation in the Summer Language Workshop at Indiana University. Her research interests include acquisition of Russian phonology by American learners and teaching pronunciation instruction.

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

REFERENCES

- Boersma, P., & Weenink, D. (2011). Praat: doing phonetics by computer (Version 5.2.43) [Computer software]. Retrieved from <http://www.fon.hum.uva.nl/praat/>
- Cebrian, J. (2000). Transferability and productivity of L1 rules in Catalan-English interlanguage. *Studies in Second Language Acquisition*, 22, 1-26.
- Dmitrieva, O., Jongman, A., & Sereno, J. (2010). Phonological neutralization by native and non-native speakers: The case of Russian final devoicing. *Journal of Phonetics*, 38, 483-92.
- Eckman, F. (1977). Markedness and the contrastive analysis hypothesis. *Language Learning*, 27, 315-30.
- Forster, K., & Forster, J. (2003). DMDX: a windows display program with millisecond accuracy. *Behavior Research Methods, Instruments & Computers*, 35, 116-124.
- Kharlamov, V. (2014). Incomplete neutralization of the voicing contrast in word-final obstruents in Russian: phonological, lexical and methodological influences. *Journal of Phonetics*, 43, 47-56.

- Mack, M. (1982). Voicing-dependent vowel duration in English and French: Monolingual and bilingual production. *Journal of the Acoustical Society of America*, 71, 173-78.
- Ohala, J. (1983). The origin of sound patterns in vocal tract constraints. In P. MacNeilage (Ed.), *The production of speech* (pp. 189-216). Berlin: Springer.
- Port, R. & Crawford, P. (1989). Incomplete neutralization and pragmatics in German. *Journal of Phonetics*, 17, 257-82.
- Prince, A., & Smolensky, P. (1993). *Optimality theory: constraint interaction in generative grammar* (Report No. 2). Rutgers University Center for Cognitive Science.
- Schwartz, B., & Sprouse, R. (1996). L2 cognitive states and the full transfer/full access model. *Second Language Research*, 12, 40-72.
- Simon, E. (2010). Phonological transfer of voicing and devoicing rules: evidence from L1 Dutch and L2 English conversational speech. *Language Sciences*, 32, 63-86.
- Smith, B., Hayes-Harb, R., Bruss, M., & Harker, A. (2009). Production and perception of voicing and devoicing in similar German and English word pairs by native speakers of German. *Journal of Phonetics*, 37, 257-275.

Zetterholm, E., & Tronnier, M. (2015). Swedish tonal word accents pronounced by Vietnamese L1-speakers. In J. Levis, R. Mohammed, M. Qian & Z. Zhou (Eds). *Proceedings of the 6th Pronunciation in Second Language Learning and Teaching Conference* (ISSN 2380-9566), Santa Barbara, CA (pp. 95-103). Ames, IA: Iowa State University.

SWEDISH TONAL WORD ACCENTS PRODUCED BY VIETNAMESE L1-SPEAKERS

[Elisabeth Zetterholm](#), Linnaeus University and Stockholm University, Sweden

[Mechtild Tronnier](#), Lund University, Sweden

This study addresses the question of whether Vietnamese L1 speakers make an adequate distinction in tonal patterns when producing two tonal accent words in their L2, Swedish. Both languages use tonal features to distinguish lexical meaning. Some previous studies suggest that a lexical tonal L1 may provide L2 learners an advantage in perceptually discriminating between different tones in another tone language, while other studies show that this may not necessarily be the case. What constitutes an adequate distinction is identified as such by native speakers/listeners of Swedish. Results revealed that no adequate distinction is made between the tone accents by the L2 speakers. However, one tonal pattern is produced more frequently and it seems to resemble one of the Swedish accents, but not the other. It may be that one of the Vietnamese tones is similar to a certain accent pattern in Swedish and is therefore recognized by the L2 speakers and transferred in tonal accent production.

INTRODUCTION

Accurate word accent processing, both perception and production, is one of the obstacles for L2 learners of Swedish. There are two word accents in Swedish based on varied tonal contours aligned with the main stressed syllable in the word. They can either be lexically distinctive or prepare the listener for the following morphological suffix. According to observations made by teachers of Swedish as a second language from sessions on pronunciation training, speakers with a tonal L1 seem to be better able to cope with this obstacle. In the study presented here, the hypothesis is examined that speakers with a lexical tonal L1 have an advantage in the production of tonal word accents in the L2. However, no adequate distinction is made between the tone accents by the L2-speakers. The identification of the two accent patterns produced by L2 speakers as compared to those produced by Swedish L1 speakers is analyzed.

Tonal Perception in L2

It has been claimed that it is an advantage if the speakers' L1 has lexical tones when it comes to perceptually discriminating among tones in another tone language (Wang et al., 2004). Speakers of non-tonal L1s, on the other hand, seem to be less sensitive to lexical tone contrast (Wayland & Guion, 2004). In cases where a speaker's L1 has pitch accents, the ability to contrast tones in a foreign tone language is comparable to the abilities of L1 speakers of a tone language (Burnham et al., 1996).

A study on perception accuracy of lexical tones (Schaefer & Darcy, 2014) suggests a typology of pitch prominence, based on findings that L1 speakers of a tone language perform best when identifying tones in L2; whereas L1 speakers of a pitch accent language perform less well,

although these speakers still perform better than speakers of an L1 with word stress and ‘intonation-only’ characteristics.

Tonal Production in L2

In a comparative study of Swedish tone accent production (Tronnier & Zetterholm, 2013), it was shown that L1 speakers of a tone language (Thai and Vietnamese) produced a less systematic distinction when compared to Somali L1 speakers, even though Somali is a tone accent language, as is Swedish. A follow up-study (Tronnier & Zetterholm, 2014) showed otherwise: that the tonal patterns used for the systematic distinction in production was not sufficient for native speakers/listeners to identify the appropriate word equally as often as those produced by L1 speakers. In this paper, the typology presented in Schaefer and Darcy (2014) will be tested beyond the production data in that the degree of adequate L2 production of tonal accents by L1 speakers of a tone language (Vietnamese) will be investigated through perception tests with L1 informants.

Tone Accents in Swedish

The two tonal word accents in Swedish are aligned to the stressed syllable in a word. The accent types are called Accent 1 (acute) and Accent 2 (grave). Two segmentally similar words can differ in accent type on the basis of variation of the pitch contour alignment, which leads to several minimal pairs distinguished only by tone accent. If lexical stress falls on the initial syllable, which is the case for the target words in this study, both Accent 1 and Accent 2 may occur and are distinctive in meaning. The word’s morphological structure, the placement of the stressed syllable and the origin of the word decide which of the two accents occurs.

Although there is variation across Swedish dialects, there is always an earlier HL-pattern (F0-fall) for Accent 1 compared to Accent 2. Very few dialects do not make any accent distinction. In Southern Swedish dialects the fall associated with Accent 1 occurs early in the stressed syllable, whereas it occurs after the stressed syllable for Accent 2 (Bruce & Gårding, 1978). Accent 2 is considered to be the marked member of the accent opposition (Riad, 1998). All speakers in this study reside in the Southern part of Sweden. Figure 1 shows one example of the two word accents in a Southern Swedish dialect.

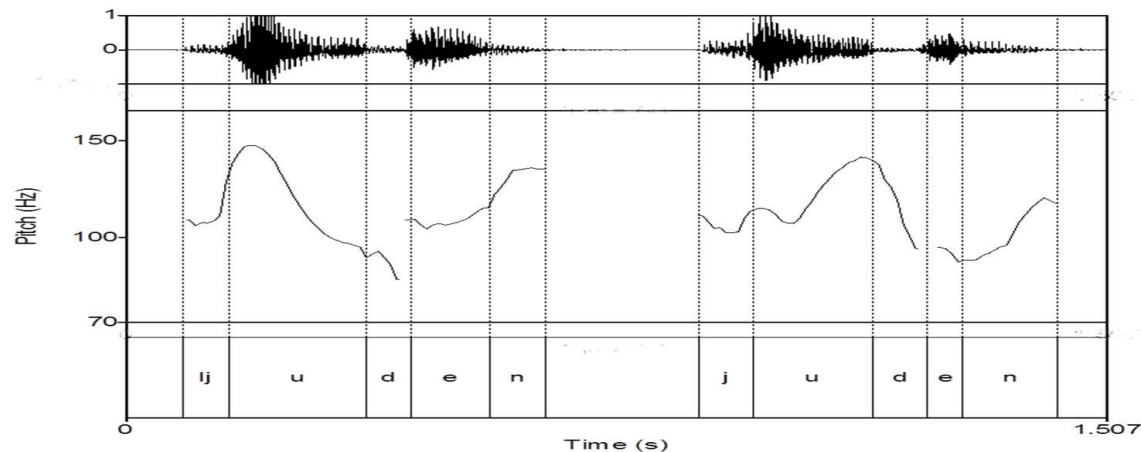


Figure 1. Realization of Accent 1 (left) and Accent 2 (right) in Southern Swedish; stress on the first syllable in both words. The words are *ljuden* (the sound) [jú:ðən] to the left and *juden* (the Jew) [jù:ðən] to the right.

Lexical Tones in Vietnamese

Vietnamese is a contour tone language with six lexical tones divided into two basic registers; high and low tones. See Figure 2. The high tones are high level, high rising and broken fall-rise (glottalized and abrupt rise). The low tones are gradual falling, low dropping and curve (gradual fall and rise) (Ingram & Nguyen, 2006). The pitch height and the direction of pitch movement are the two primary dimensions to contrast lexical items and words. In addition, tones are distinguished by voice quality, intensity, and duration (Nguyen & Ingram, 2005). However, direction of pitch movement, pitch height and voice quality are more important features than other tonal dimensions such as duration and intensity in tone recognition.

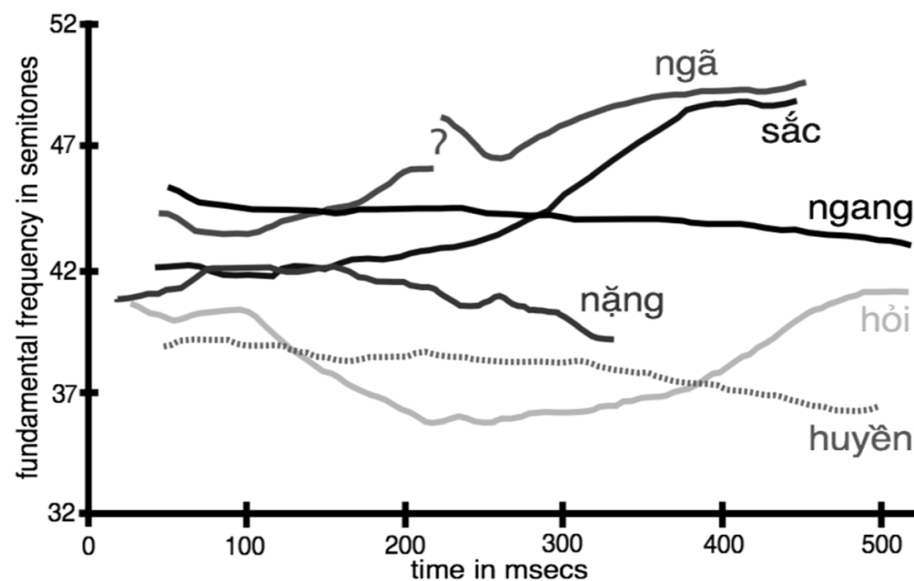


Figure 2. Vietnamese tones (adopted from Nguyen & Edmondson, 1998).

THE INVESTIGATION

The present study is concerned with the acquisition of the L2 tone accent distinction. The accuracy of the two accent patterns produced by Swedish L2 speakers with Vietnamese as their L1 is examined by means of identification tests, carried out by listeners with Swedish L1. The Swedish L2 speakers investigated here did not receive any specific training in the Swedish tone accents when studying Swedish as a second language. Therefore, the knowledge they have about the distinction between the two word accents has developed when practicing Swedish in their daily lives. It is, however, not clear whether the L2 speakers are aware of the fact that tonal accent distinction occurs in Swedish. An acoustic analysis of the Vietnamese speakers' production of the two word accents shows a mixed pattern. For some examples there is no distinction, but in most cases a pattern similar to Accent 2 is preferred (Tronnier & Zetterholm, 2013). The identification rate of the stimuli produced by Vietnamese L1 speakers is compared to the results of the identification of a matching set of stimuli produced by speakers with Swedish as their L1. In addition, listeners were asked to rank the degree of difficulty encountered when deciding which tone accent they heard during the identification test.

Recordings, test material and listeners

Recordings of two native speakers of Swedish and two native speakers of Northern Vietnamese with Swedish as their L2 were used for this study. The recordings consisted of read speech and the sentences were prepared with four minimal pairs contrasted by word accents (see Table 1). The words are quite common Swedish words but were chosen primarily because they are minimal pairs. The eight target words produced by each speaker were carefully cut out from the recordings and the resulting 32 stimuli were randomized and presented to 20 listeners with Swedish as their L1. The listeners' ages ranged from 27 to 74 and the average age was 42. About three-fourths of the listeners indicated that they speak a dialect from the Southern or the Southwest part of Sweden. They all lived in the south of Sweden, and they all speak a dialect with a distinction between the two word accents. The influence of dialectal variation among the listeners, however, will not be taken into further consideration in this study. The test was constructed in the learning platform MyMoodle. All listeners carried out the test on their own computers, using local loudspeakers or headsets. After listening to a stimulus, they had to check one of the two listed words, a forced choice test. They could listen to the stimulus as many times as they wanted, but could not return to earlier items in the test. They also had to express how difficult it was to discriminate the accent on a scale of 1-5, where 1 represented a very easy decision and 5 a very difficult decision.

Table 2

Target words and their morphological structure: all words are stressed on the first syllable, which is where the difference in tonal accent is also found.

Accent 1	Accent 2
<i>fäster</i> (attach) [fɛ̂stər], verb, 3 rd pers.sing.	<i>fester</i> (parties) [fɛ̂stər], noun, common gender, pl.indef.
<i>Oskar</i> (name for a boy) [óskar]	<i>åskar</i> (thunderstorm) [òskar], verb, 3 rd pers.sing.
<i>stegen</i> (steps) [sté:gən], noun, neutrum, pl.def.	<i>stegen</i> (ladder) [stè:gən], noun, common gender, sing.def.
<i>tecken</i> (sign) [tékən], noun, neutrum, pl.def.	<i>täcken</i> (bed cover) [têkən], noun, neutrum, sing.def.

Data Analysis

The number of correct identifications and the experienced degree of difficulty of identification of the target words in relationship to the speakers' L1 obtained from the listeners were statistically analyzed using t-tests for independent samples. In addition, the influence of the individual tone accents was taken into consideration.

RESULTS

Identification

The diagrams in Figures 3-6 give an overview of the general results of this study and also provide detailed information and more specific results. In the diagrams, the mean identification rate (y-axis) of the eight target words (x-axis) across the 20 listeners is presented, where the filled part (lower) of the columns represents the correct answers and the barred part (upper) the inaccurate answers. Each diagram represents the results for one of the four speakers, who produced stimuli for this study. Furthermore, the columns are organized such that the members of a minimal pair are placed next to each other, target words carrying Accent 1 are labeled with '1', and those target words carrying Accent 2 are labeled with '2'.

Correct identification was very much dependent on the speakers' L1 in that the stimuli produced by L1 speakers of Swedish were identified significantly more often, although not in every case ($p < .005$). For either L1, no difference in identification was found between the stimuli produced by the two speakers (Swedish L1: $p > 0.5$ and Vietnamese L1: $p > 0.1$). None of the accents is more often correctly identified for the stimuli produced by the Swedish L1 speakers ($p > 0.1$). For the L2 speakers, however, Accent 2-words were more often correctly identified than Accent 1-words ($p < 0.01$). Comparing the identification rate between the two groups of speakers, Accent 1-words were more frequently identified when produced by L1 speakers ($p < 0.001$) than the Accent 2-words ($p > 0.5$). Accent 2-words were generally more successfully identified for L2 speakers than

their Accent 1-words, although with slightly less accuracy than Accent 2-words produced by L1-speakers.

Difficulty

Tone accent identification was significantly more difficult for the stimuli produced by the Vietnamese speakers ($p < 0.005$). Within each group of L1 speakers, neither of the accents was more difficult to identify than the other (for Swedish L1: $p > 0.05$ and for Vietnamese L1: $p > 0.1$), although the results indicate that stimuli produced by L2 speakers were generally more difficult to identify. However, the confidence for correct identification varied between different types of word accents for the different L1s in that the marked Accent 2 was experienced to be equally difficult to identify for the stimuli regardless of the producer's L1 ($p > 0.1$). This was not the case for Accent 1-words. For unmarked Accent 1-words, stimuli produced by the Swedish L1 speakers were much easier to identify than those produced by Vietnamese L1 speakers ($p < 0.005$).

Correct Identification vs. Experienced Difficulty

There is a general correlation between the number of correctly identified stimuli and the degree of difficulty experienced, in that the stimuli produced by Swedish L1 speakers were identified correctly significantly more often and were experienced as being significantly easier to identify. Identification of the individual accent was equally accurate and equally difficult in the case of Swedish L1 speakers. In the case of L2 speakers, however, Accent 2-words were identified correctly more frequently than Accent 1-words, although they were experienced to be equally difficult. Accent 2-words were correctly identified to the same extent and rated as equally difficult for both groups of speakers, whereas Accent 1-words were both less correctly identified and also judged to be more difficult when produced by L2 speakers.

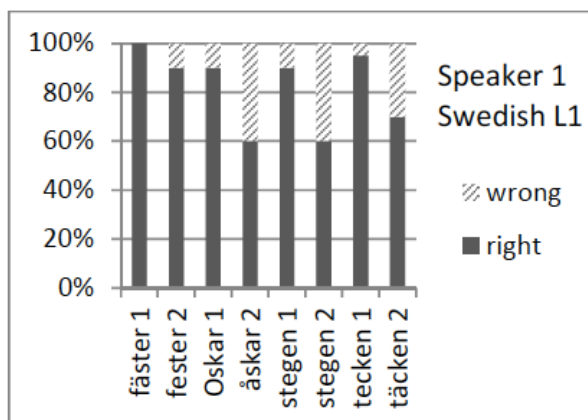


Figure 3. Mean identification rate (%) of the eight target words across the 20 listeners for stimuli produced by one of the speakers with Swedish L1, Speaker 1.

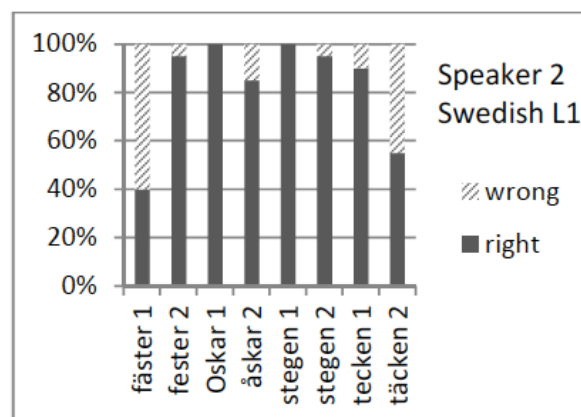


Figure 4. Mean identification rate (%) of the eight target words across the 20 listeners for stimuli produced by one of the speakers with Swedish L1, Speaker 2.

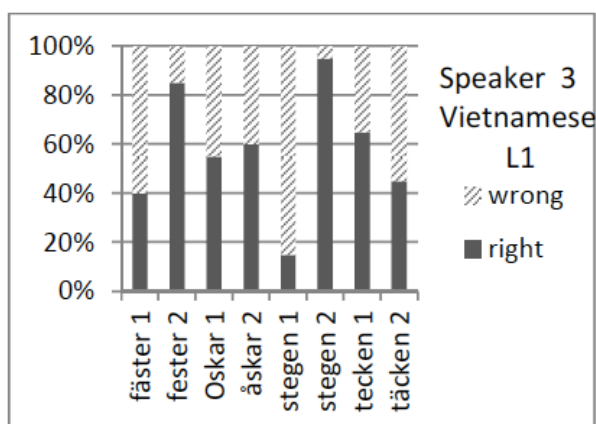


Figure 5. Mean identification rate (%) of the eight target words across the 20 listeners for stimuli produced by one of the speakers with Vietnamese L1, Speaker 3.

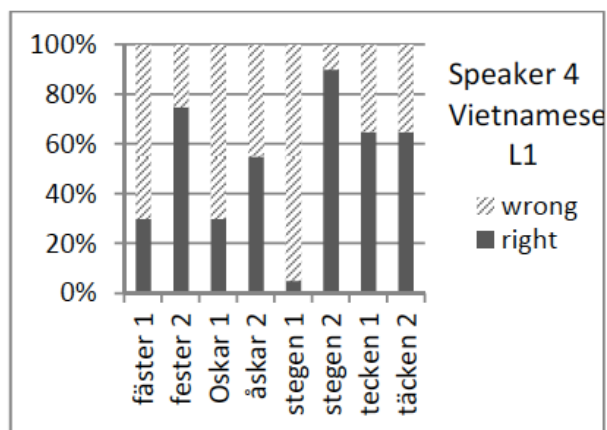


Figure 6. Mean identification rate (%) of the eight target words across the 20 listeners for stimuli produced by one of the speakers with Vietnamese L1, Speaker 4.

DISCUSSION

The tone accents were generally identified with less accuracy and were perceived to be more difficult to identify for the stimuli produced by L2 speakers. This leads to the conclusion that the tonal patterns produced to distinguish the tonal accents in L2 Swedish are not sufficiently similar to what is expected by the listeners, who are L1-speakers of Swedish. This confirms the results in Tronnier & Zetterholm (2013), which showed from an analysis of produced pitch patterns that no systematic distinction between Accent 1-words and Accent 2-words was made. Placing the current results within a typological framework similar to the one suggested in Schaefer and Darcy (2014), i.e., how L1 speakers of languages with different tone typologies process tonal characteristics in other languages of the same or different level of tonal typology, has to be investigated more systematically. Such an investigation should include comparisons between results from similar studies of L1 speakers of other languages with varied tonal typology, such as those found in Tronnier and Zetterholm (2014).

There is, however, more to say about how the individual word accents were dealt with. It is clear that most of the tonal patterns produced by the Vietnamese L1 speakers were perceived as Accent 2-patterns, and in most cases they were correctly identified as Accent 2-words, but in many cases they were misidentified as Accent 2-words that should have been Accent 1. Furthermore, Accent 2-words were rated as being easier to identify. It is therefore plausible that L2 speakers seem to be comfortable producing a tonal pattern which is closely related and acceptable for an Accent 2-contour, although it is used for the wrong words at times, i.e., the tonal pattern was overgeneralized. The question that arises here is whether one of the tones in Vietnamese resembles the Swedish Accent 2-pattern very closely. A systematic investigation is needed to relate and compare the tones in Vietnamese with the tonal patterns in Swedish. This is especially necessary for the tonal pattern of Accent 2.

It has been claimed that the Accent 2 is the marked accent in Swedish (Riad, 1998). This also means that the contour assigned to Accent 2 allows less tonal variation than it does in the unmarked case, Accent 1. In that respect, listeners indicated that some contours fit within the restricted tonal framework for Accent 2, a finding that also supports the idea that a corresponding tonal pattern adequate for Accent 2 in Swedish may exist in Vietnamese. So far, no decision can be made as to which tone in Vietnamese could be the appropriate one (see Figure 2). In contrast, it has been shown in a previous study (Tronnier & Zetterholm, 2014), that L2-speakers, with an L1 other than Vietnamese, produced tonal patterns more accurately for Accent 1-words, a finding which was interpreted as being in agreement with the hypothesis that Accent 1 is the unmarked accent.

In summary, the assignment of tonal contours in L2 Swedish produced by the L1 speakers of Vietnamese – a tone language – was not adequate. However, tonal patterns acceptable as Accent 2-contours were produced more frequently and were identified as such, although they were also incorrectly applied to Accent 1-words. In conclusion, the L1 speakers of the tone language Vietnamese do not generally produce appropriate tonal patterns or distinguish them in an appropriate way in L2 Swedish. They often produce one, probably familiar, tone contour that is acceptable for one of the Swedish tonal accents, namely Accent 2. However, they also produce this pattern for words where it is not acceptable.

ACKNOWLEDGMENTS

The current research was partly funded by Linnaeus University.

ABOUT THE AUTHORS

Elisabeth Zetterholm (elisabeth.zetterholm@isd.su.se) is a senior lecturer and researcher in Swedish as a second language at Linnaeus University and Stockholm University, Sweden. Her current research interests focus on pronunciation in second language acquisition, comprising segmental and prosodic issues. When teaching, most of her students are prospective teachers of Swedish as a second language and therefore she wants to develop the teaching methodology. She received her PhD in phonetics at Lund University, Sweden (2003). Her thesis and postdoctoral research were about voice imitation, with a focus on speaker identification and individual features in voice and speech.

Mechtild Tronnier (mechtild.tronnier@ling.lu.se) is a senior lecturer and researcher in phonetics at Lund University, Sweden. Her current research interest focuses on pronunciation in second language acquisition, comprising segmental and prosodic issues. When lecturing, her students are enrolled in diverse programs and range from prospective teachers of Swedish as a second language, speech pathology students to students in linguistics on both undergraduate and graduate levels. She received her PhD in phonetics at Lund University, Sweden (1998). Her thesis work was concerned with nasals and nasality with a focus on Japanese.

REFERENCES

- Bruce, G. & Gårding, E. (1978). A prosodic typology for Swedish dialects. In E. Gårding, G. Bruce & R. Bannert (Eds.) *Nordic prosody: Papers from a symposium* (pp. 219-228). Lund: Lund University, Department of Linguistics.
- Burnham, D., Francis, E., Webster, D., Luksaneeyanawin, S., Attapaiboon, C., Lacerda, F. & Keller, P. (1996). Perception of lexical tone across languages: Evidence for a linguistic mode of processing. *Proceedings of the Fourth International Conference on Spoken Language processing (ICSLP96, pp. 2514-251), Philadelphia, USA.*
- Ingram, J. & Nguyen, T. T. A. (2006). Stress, tone and word prosody in Vietnamese compounds. *Proceedings of the 11th Australian International Conference on Speech Science & Technology*, University of Auckland, New Zealand.
- Nguyen, T.T.A. & Ingram, J. (2005). Vietnamese acquisition of English word stress. *TESOL Quarterly*, 39, 309-319. ,
- Nguyen, V. L. & Edmondson, J. A. (1998). Tones and voice quality in modern northern Vietnamese: Instrumental case studies. *Mon-Khmer Studies*, 28, 1-18.
- Riad, T. (1998). Towards a Scandinavian accent typology. In W. Kehrein & R. Wiese (Eds.) *Phonology and Morphology of the Germanic Languages* (pp. 77-109). Berlin: Walter de Gruyter.

- Schaefer, V. & Darcy, I. (2014). Pitch prominence matters: Perception of Thai tones by Seoul Korean and Kyungsang Korean speakers. *Concordia Working Papers in Applied Linguistics*, 5, 597-611.
- Tronnier, M. & Zetterholm, E. (2013). Tendencies of Swedish word accent production by L2-speakers with tonal and non-tonal L1. In E. L. Asu & P. Lippus (Eds.) *Proceedings of Nordic prosody XI* (pp. 391-400). Tartu, Estonia.
- Tronnier, M. & Zetterholm, E. (2014). Appropriate tone accent production in L2 Swedish by L1 speakers of Somali? *Concordia Working Papers in Applied Linguistics*, 5, 722-736.
- Wang, Y., Behne, D. M. Jongman, A. & Serano, J.A. (2004). The role of linguistic experience in the hemispheric processing of lexical tones. *Applied Linguistics*, 25, 449-466.
- Wayland, R. P. & Guion, S. G. (2004). Training English and Chinese listeners to perceive Thai tones: A preliminary report. *Language Learning*, 54, 681-712.

Inceoglu, S. (2015). Audiovisual and auditory-only perceptual training: Effects on the pronunciation of French nasal vowels. In J. Levis, R. Mohammed, M. Qian & Z. Zhou (Eds). *Proceedings of the 6th Pronunciation in Second Language Learning and Teaching Conference* (ISSN 2380-9566), Santa Barbara, CA (pp. 104-114). Ames, IA: Iowa State University.

AUDIOVISUAL AND AUDITORY-ONLY PERCEPTUAL TRAINING: EFFECTS ON THE PRONUNCIATION OF FRENCH NASAL VOWELS

Solène Inceoglu, Rochester Institute of Technology

Previous studies have shown that gains made during perceptual training can be transferred to gains in production and that audiovisual (AV) perceptual training often leads to greater improvement in production than auditory-only (A-only) training (Hardison, 2003; Hazan, Sennema, Iba, & Faulkner, 2005). This study investigated whether perceptual training on the three French nasal vowels led to improvement in the production of these vowels, and whether improvement was greater with AV perceptual training as opposed to A-only training. The productions of 60 American-English intermediate learners of French were recorded at pretest and posttest. The stimuli consisted of 108 CVC words in various consonantal contexts and the L2 learners' productions were judged by two native French listeners in two rating tasks: a forced-choice identification rating task and a quality rating task. Results showed that both training groups—but not the control group—significantly improved from the pretest to the posttest, but that the production of the AV training group improved significantly more than the production of the A-only training group. Furthermore, the two types of analysis used to assess the production of the L2 learners revealed differences that have implications for research methodology and assessment.

INTRODUCTION

There is a general consensus that most individuals who learn a second language (L2) as adults speak it with a foreign accent (e.g., Derwing & Munro, 1997; Flege, Munro, & MacKay, 1995; Major, 2001). This fact has led to proposals for several model of speech perception. The Speech Learning Model (Flege, 1992, 1995), for instance, claims that the similarity between L1 and L2 sounds makes it difficult for learners to perceive phonetic differences and to create new categories for L2 sounds. The Similarity Differential Rate Hypothesis (Major & Kim, 1996) goes further and argues that dissimilar sounds between an L1 and L2 are acquired faster than similar sounds. This crosslinguistic influence in perception is also believed to be reflected in production, and numerous studies have shown that the degrees of accuracy in perceiving and producing L2 phones are related (e.g., Flege, Bohn, & Jang, 1997; Flege, 1988). Of interest to second language acquisition (SLA) researchers and teachers is the question of how novel L2 phoneme categories are acquired and the extent to which instruction or training can improve production. A wealth of empirical studies have provided evidence that perceptual auditory training can be transferred to improvement in production even when no production tasks are involved during training (e.g., Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999; Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Lambacher, Martens, Kakehi, Marasinghe, & Molholt, 2005; Lopez-Soto & Kewley-Port, 2009; Wang, Jongman, & Sereno, 2003).

A number of different methods have been used assess L2 learners' speech production. Among the tasks commonly used are forced-choice identification tasks (Bradlow et al., 1997; Hazan et al., 2005), identification tasks (Lambacher et al., 2005), quality rating tasks (Hardison, 2003;

Hazan et al., 2005), acoustic analyses (Lambacher et al., 2005; Wang et al., 2003) and pretest-posttest paired comparison (Bradlow et al., 1997). In addition, while some studies use only one type of rating task (e.g., Hardison, 2003; Lopez-Soto & Kewley-Port, 2009; Wang et al., 2003), others combine several tasks (e.g., Bradlow et al., 1997; Hazan et al., 2005; Lambacher et al., 2005).

Audiovisual Speech Perception

The large majority of L2 speech studies have focused on one source of input—the auditory signal—despite the fact that speakers rely on both auditory and visual information in face-to-face communication, making communication a multimodal experience (Pisoni & Remez, 2005; Rosenblum, 2005). The ability to benefit from both the auditory and visual modality in perceiving and interpreting speech has been demonstrated with native speakers in studies investigating L1 speech comprehension (e.g., Arnold & Hill, 2001; Reisberg, McLean, & Goldfield, 1987; Summerfield, 1979), speech intelligibility (e.g., Benoît, Mohamadi, & Kandel, 1994; Sumby & Pollack, 1954), and language discrimination (e.g., Ronquest, Levi, & Pisoni, 2010; Soto-Faraco et al., 2007). For instance, Arnold and Hill (2001) found that comprehension was better in audiovisual conditions than in auditory-only conditions when messages were presented in various ways: with accented speech, in a L2 that participants were fluent in, or with complex semantic and syntactic structures. In addition, although non-native speakers have been found to be less efficient at using visual information than native speakers (e.g., Hazan et al., 2005, 2006; Ortega-Llebaria et al., 2001), the superiority of audiovisual information over auditory-only information has also been supported in studies looking at L2 speech perception (e.g., Erdener & Burnham, 2005; Hardison, 2003; Hazan et al., 2006, 2005; Hirata & Kelly, 2010; Kluge, Reis, Nobre-Oliveira, & Bettoni-Techio, 2009; Wang, Behne, & Jiang, 2008).

Despite a growing number of studies pointing to the facilitative aspects of audiovisual speech information in attending to L2 speech, only a couple of studies have examined the relationship between audiovisual speech perception and production (Erdener & Burnham, 2005) or looked at the effect of audiovisual training on the production of L2 phonemes (Hardison, 2003; Hazan et al., 2005). Because of the relative paucity of published studies, it is premature to draw firm conclusions. Nonetheless, two studies investigating the production of the English /r-l/ contrast by Japanese speakers reported that audiovisual perceptual training led to greater improvement in production than auditory-only training (Hazan et al., 2005). Production performance was also found to be superior when visual information (i.e., the face of the speaker) was present in a study where Australian English and Turkish speakers were asked to repeat words produced in Spanish and Irish (Erdener & Burnham, 2005).

Still, much work remains to be done regarding the effect of L2 speech perception on speech production. Accordingly, this study contributes to the ongoing research on audiovisual speech perception by exploring whether perceptual training helps improving the production of the three (Standard) French nasal vowels. The rationale for targeting nasal vowels is that their difference is visually salient, ranging from the hyper-rounded [ɔ̃] to the rounded [ɑ̃] and unrounded [ɛ̃], and they are often problematic for non-native speakers.

Research Questions

The general research question addressed in this article was whether participants receiving audiovisual (AV) perceptual training improve their pronunciation of the French nasal vowels more than participants receiving audio-only (A-only) training. In order to have a better understanding of assessment methods and rating tasks, the research question was divided into two sub-questions examining participants' pronunciation when (a) rated by native speakers in a quality rating task, and (b) rated by native speakers in a forced-choice identification task.

METHODS

Participants

Sixty participants (age 18-24, mean = 20; 43 females) were recruited from intermediate French language courses at a large Midwestern university and were randomly assigned to one of the three following groups: AV training, A-only training, and a control group. All the participants were native speakers of American English who reported good vision, no hearing disorder, and no background in lipreading and phonetics.

Stimuli and Procedure

The participants were tested at pretest and posttest individually in a quiet room. A delayed repetition task was used to elicit participants' production of 108 monosyllabic #CVC# words where the vowel was either [ã], [ɔ̃], or [ɛ̃]. The initial consonant was one of the following: [p-t-k-b-d-g-s-z-f-v-ʒ-j] to take into consideration the articulation of vowels in different consonantal contexts. Participants heard the stimuli followed by one of five version of a prompt asking them, in French, to repeat the word. For example, they would hear the stimulus "ponse" [pɔ̃s] followed by "*répète le mot s'il te plaît*" (repeat the word, please). The reason for having a prompt between the stimuli and the participants' repetition of the stimuli was to prevent direct imitation from sensory memory.

Two rating tasks were used to assess participants' oral production. Two native listeners of French rated the pretest and posttest productions of the sixty participants in a forced-choice identification task and a quality rating task. For both tasks, raters were asked to focus on the accuracy of the vowels production and ignore the production of the consonants. In the forced-choice identification task, the raters heard a stimulus produced by L2 participants and were asked to choose which of the three nasal vowel the L2 participants had produced. The task was not timed, and raters had the option to listen to the production a second time, if needed. The two native listeners rated 12,960 tokens (108 stimuli × 2 tests (pre and post) × 60 participants) and the inter-rater reliability was 88.4%, which is considered acceptable. Discrepancies were rated by a third native rater to ensure that all responses used for the analysis were agreed on by at least two native speakers of French. In the quality rating task, raters were presented with a target word on a computer screen, listened to a participant's production of the word and rated the production on a scale from 1 (bad) to 7 (excellent). Participants' pretest and posttest productions of each single word were presented in one block, resulting in a total of 108 blocks—one for each word. The average between the ratings of the two native listeners was calculated and used for the

quality rating task. The order of the production tokens was randomized across raters and the ratings were completed within a period of one month.

Perceptual Training

Between the production pre- and post- tests described above, the two training groups received six 30-minute sessions of either AV (hearing a speaker while seeing her face) or A-only (just hearing the speaker) high-variability phonetic perceptual training. Participants sat in front of a computer screen and were presented with a #CVC# stimulus similar to those used for the production task. They then had 4000 milliseconds to click on one of the three options (e.g., nasal vowel “an”, “on”, or “un”) on the screen before receiving feedback. Participants were presented with 178 CVC stimuli randomized across participants and training sessions.

RESULTS

Forced-choice Identification Rating Task

The mean production accuracy scores at pretest and posttest are illustrated in Figure 1. A repeated-measures ANOVA with Time as the repeated measure and Group as between-subject factor was conducted to measure the effect of perceptual training on production accuracy. Results indicated a significant effect of Time, $F(1, 5989) = 111.17, p < .001$, as well as a significant interaction between Group and Time, $F(2, 5989) = 24.16, p < .001$. This interaction was analyzed with separate t-tests (adjusted alpha level of .016) which showed that there were significant differences between the pretest scores ($M = .63, SD = .48$) and posttest scores ($M = .76, SD = .42$) of the AV group, $t(1833) = 10.20, p < .001$, as well as between the pretest scores ($M = .66, SD = .47$) and posttest scores ($M = .73, SD = .44$) of the A-only group, $t(2110) = 6.38, p < .001$. There was, however, no significant differences between the pretest scores ($M = .68, SD = .46$) and posttest scores ($M = .70, SD = .46$) of the control group, $t(2046) = 1.14, p = .254$, providing evidence that improvement in production accuracy was related to perceptual training. A one-way ANOVA with Training as the between-subjects factor and Gain change as the dependent variable revealed that the improvement in production accuracy of the AV training group was significantly greater than the improvement of the A-only training group, $F(1, 3944) = 11.95, p = .001$, thus confirming the superiority of AV perceptual training over A-only training as far as transfer to production is concerned.

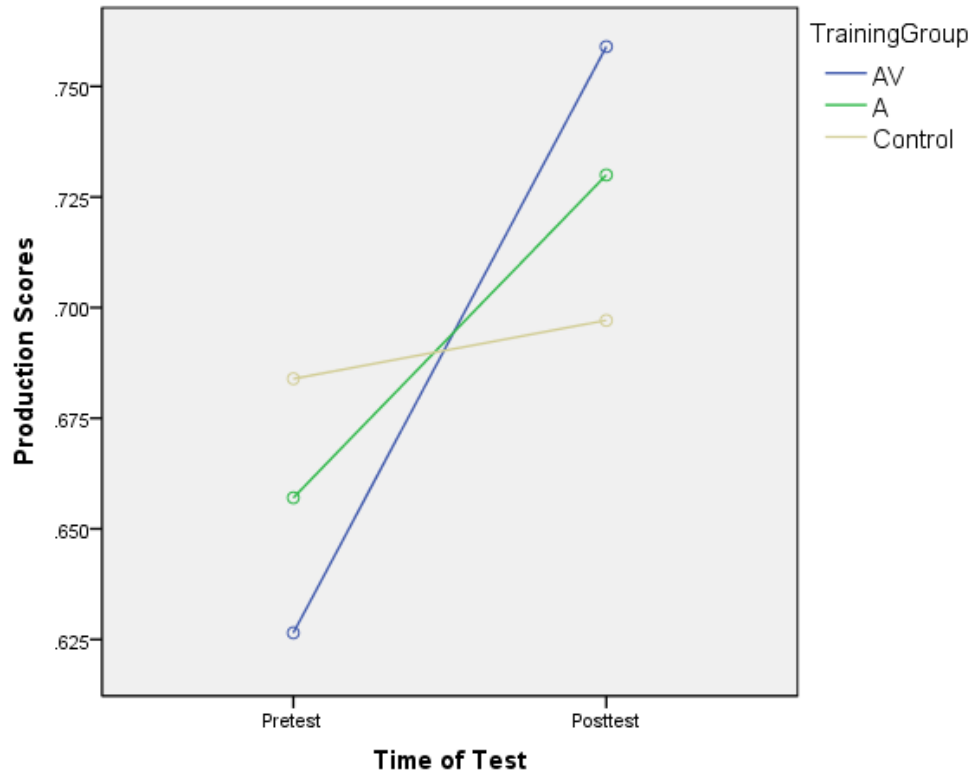


Figure 1. Accurate production at pretest and posttest as rated by native speakers in a forced-choice identification task (maximum score 1).

Further analyses were conducted to compare production in each of the three vowels (Table 1). Results indicate that AV training led to significantly greater improvement in the pronunciation of [ɔ̃] than A-only training ($p = .014$) and that the two types of training led to greater improvement than no training. The difference in the improvement of [ã] was not statistically different between the two training groups ($p = .59$) or between the A-only training and control groups ($p = .225$), but the AV training group improved more than the control group ($p = .028$). Finally for [ẽ], the AV training group improved significantly more than the A-only training ($p = .046$) and the control group ($p < .001$), but the A-only training did not improve more than the control group ($p = .185$).

Table 1

Percentage of correct identification by native speakers of vowels produced by L2 speakers at pretest and posttest

		[ʃ]	[ã]	[ê]
AV training	Pretest	58.75	54.99	74.18
	Posttest	77.22	58.83	90.12
A-only training	Pretest	60.77	56.18	78.88
	Posttest	72.01	58.23	88.79
Control	Pretest	63.54	60.96	80.79
	Posttest	65.64	56.72	86.80

Quality Rating Task

Results of the quality rating tasks are illustrated in Figure 2. Repeated-measures ANOVAS showed that all groups significantly improved: $F_{AV}(1, 1833) = 152.73, p < .001$; $F_A(1, 2157) = 43.00, p < .001$; $F_C(1, 2048) = 33.91, p < .001$. A one-way ANOVA with Group as the between-subjects factor and Gain change as the dependent variable revealed a significant effect of Group, $F(2, 6040) = 7.70, p < .001$ and a Bonferroni post-hoc test identified that the AV group improved significantly more than the A-only group ($p = .002$) and the control group ($p = .002$). However, and contrary to the overall results of the forced-choice identification task, the A-only group did not appear to have improved more than the control group did ($p = .10$).

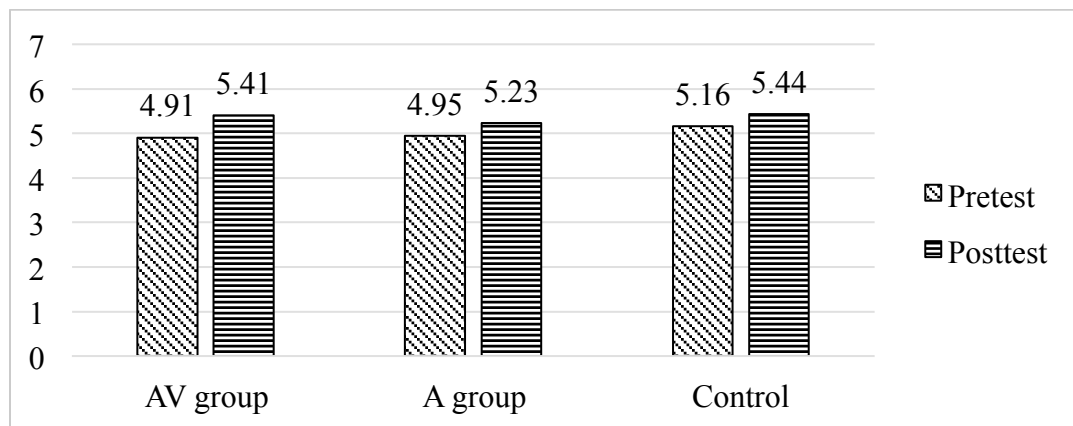


Figure 2. Percentage of accurate production at pretest and posttest as rated by native speakers in a quality rating task.

Detailed analyses for each vowel (Table 2) showed that the difference in the improvement of the pronunciation of [ʃ] by the two training groups was not significant ($p = .094$), and that the two training groups improved significantly more than the control group (AV: $p < .001$, A: $p = .029$). As for [ã], not only did the AV training group improve significantly more than the A-only training group ($p = .022$), but the performance of the latter actually decreased

from pretest to posttest. Furthermore, the difference in pronunciation improvement between the control group and the two other groups did not reach significance (AV: $p = .352$, A: $p = .405$). Finally, there was no significant effect of Group, $F(2, 2013) = 0.70$, $p = .494$ for the improvement of [ɛ̃], indicating that all groups improved in a similar way.

Table 2

Means of production rating (7-point scale) for each vowel at the pretest and posttest

		[ɔ̃]	[ɑ̃]	[ɛ̃]
AV training	Pretest	4.98	4.69	5.07
	Posttest	5.62	4.9	5.7
A-only training	Pretest	4.94	4.82	5.09
	Posttest	5.39	4.74	5.58
Control	Pretest	5.29	4.85	5.35
	Posttest	5.49	4.91	5.93

DISCUSSION

The primary goal of this study was to compare the effect of audiovisual and audio-only perceptual training on the production of French nasal vowels by intermediate American-English learners of French. Results showed that, overall, after six sessions of perceptual training, the pronunciation of the learners in the two training groups improved significantly more than the pronunciation of the control group. These results are comparable to the perceptual training studies mentioned earlier and demonstrate that improvement in pronunciation can be achieved using a simple perceptual training task. More importantly, the results of this study demonstrated that AV perceptual training was more efficient than auditory-only training, and that participants in the AV training group benefited from the visual information they received during training and that they transferred their knowledge onto production.

The results also showed that the effects of perceptual training were stronger for the vowel [ɛ̃], which was about 90% accurately produced by both training groups at posttest. Perceptual training did not, however, seem to be as beneficial for [ɑ̃], which showed little improvement from pretest to posttest. This may be due to the lack of visual saliency of that vowel, which is neither hyperrounded, like [ɔ̃] nor unrounded like [ɛ̃].

The secondary goal of this study was to compare two types of rating tasks to get a better understanding of the data. The two tasks used showed a general similar pattern, namely that AV training was superior to A-only training, but also revealed some differences. It has previously been noted that differences in results across studies may be due to the use of different testing materials and rating procedures (Flege, MacKay, & Meador, 1999), settings, language background, and language proficiency. In the current case, the differences in results can only be explained by the rating tasks used since the participants and raters remained the same. The forced-choice identification task provided a stronger base for the argument in favor of AV training: for all vowels, the AV trainees improved significantly more than the control group, but

the AV training did not seem to be more effective than auditory-only training only for the vowel [ã], probably again because of the vowel's lack of visual saliency. On the other hand, the quality rating task led to a less straightforward picture. Results indicated that in terms of vowel quality, the AV trainees' pronunciation of [ã] improved significantly more than the pronunciation of the A-only trainees, but that AV training did not help to improve the pronunciation of [õ] and [ẽ] more than A-only training did. Finally, results of the quality task revealed that no group was better than the other two groups at improving the pronunciation of the vowel [ẽ].

Possible reasons for the discrepancy in the results is that there is a risk of score inflations when using a quality rating task. The raters rarely used the lower number of the seven-point Likert scale, and although score inflation was consistent throughout the rating task and therefore not problematic for the analysis, it became more of an issue when comparing the results of the quality rating task to those of the forced-choice identification tasks which did not leave any option for inflation. On the other hand, the forced-choice identification task was not exempt from drawbacks. First, the total accuracy scores in this study might be lower than if only two sounds had been contrasted, which is often the case in the previously cited studies, as it was somewhat cognitively more demanding as the raters have to focus on three sounds. In addition, the fact that there was no possibility to select a "none of these sounds" option forced the raters to select one option by default even if the sound produced by a L2 participant did not correspond to any of the three vowels.

In conclusion, this study contributes to the growing literature showing that AV perceptual training not only leads to improvement in speech production, but also leads to greater improvement than auditory-only training (Hardison, 2003; Hazan et al., 2005). This has strong practical implications for computer-assisted language learning and should be further investigated. In addition, this study raised methodological concerns about the assessment of speech production by showing that different results can be obtained from different types of rating.

ACKNOWLEDGMENTS

The author would like to thank Drs. Debra Hardison, Aline Godfroid, Shawn Loewen, and Anne Violin-Wigent for their advice during this research project. This research was supported by a *Language Learning Dissertation Grant*, and both a *Dissertation Completion Fellowship* from the College of Arts and Letters and a *Research Enhancement Award* from the Graduate School at Michigan State University.

ABOUT THE AUTHOR

Solène Inceoglu is Assistant Professor in the Department of Modern Languages and Cultures at Rochester Institute of Technology (Rochester, NY) where she teaches French, French Linguistics, and German Culture. She received her Ph.D. in Second Language Studies from Michigan State University in 2014 where she taught courses on SLA, Language Teaching Method, and French Phonetics. Her research focuses on SLA, L2 speech perception/production, audiovisual L2 speech processing, co-speech gesture, and psycholinguistics. She has presented at AAAL, SLRF, Eurosla, New Sounds, ASA, AILA and CALICO. Contact information: scigsl@rit.edu

REFERENCES

- Arnold, P., & Hill, F. (2001). Bisensory augmentation: A speechreading advantage when speech is clearly audible and intact. *British Journal of Psychology*, 92(2), 339–355.
- Benoît, C., Mohamadi, T., & Kandel, S. (1994). Effects of phonetic context on audio-visual intelligibility of French. *Journal of Speech and Hearing Research*, 37(5), 1195–203.
- Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception and Psychophysics*, 61(5), 977–985.
- Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. I. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America*, 101(4), 2299–2310.
- Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility. *Studies in Second Language Acquisition*, 20, 1–16.
- Erdener, D., & Burnham, D. (2005). The role of audiovisual speech and orthographic information in nonnative speech production. *Language Learning*, 55(2), 191–228.
- Flege, J. E. (1988). The production and perception of foreign language speech sounds. *Human Communication and Its Disorders: A Review*, 2, 224–401.
- Flege, J. E. (1992). The intelligibility of English vowels spoken by British and Dutch talkers. In R. D. Kent (Ed.), *Intelligibility in speech disorders: Theory, measurement, and management* (pp. 157–232). Amsterdam: John Benjamins.
- 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.
- Flege, J. E., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, 25(4), 437–470.
- Flege, J. E., MacKay, I. R. A., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *The Journal of the Acoustical Society of America*, 106(5), 2973–2987.
- Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995). Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America*, 97(5), 3125–3134.
- Hardison, D. M. (2003). Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics*, 24(4), 495–522.

- Hazan, V., Sennema, A., Faulkner, A., Ortega-Llebaria, M., Iba, M., & Chung, H. (2006). The use of visual cues in the perception of non-native consonant contrasts. *The Journal of the Acoustical Society of America*, 119(3), 1740–1751.
- Hazan, V., Sennema, A., Iba, M., & Faulkner, A. (2005). Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Communication*, 47(3), 360–378.
- Hirata, Y., & Kelly, S. D. (2010). Effects of lips and hands on auditory learning of second-language speech sounds. *Journal of Speech, Language, and Hearing Research*, 53, 298–310.
- Kluge, D. C., Reis, M. S., Nobre-Oliveira, D., & Bettoni-Techio, M. (2009). The use of visual cues in the perception of English syllable-final nasals by Brazilian EFL learners. In M. A. Watkins, A. S. Rauber, & B. O. Baptista (Eds.), *Recent research in second language phonetics/phonology: Perception and production*. (pp. 141–153). Cambridge Scholars Publishing.
- Lambacher, S. G., Martens, W. L., Kakehi, K., Marasinghe, C. a., & Molholt, G. (2005). The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Applied Psycholinguistics*, 26(2), 227–247.
- Lopez-Soto, T., & Kewley-Port, D. (2009). Relation of perception training to production of codas in English as a second language. *Proceedings of Meetings on Acoustics*, 6(1), 1–15.
- Major, R. (2001). *Foreign accent: The ontogeny and phylogeny of second language phonology*. Mahwah, NJ: Erlbaum.
- Major, R., & Kim, E. (1996). The similarity differential rate hypothesis. *Language Learning*, 46, 465–496.
- Pisoni, D. B., & Remez, R. E. (2005). *The handbook of speech perception*. (D. B. Pisoni & R. E. Remez, Eds.). Oxford, UK: Blackwell Publishing.
- Reisberg, D., McLean, J., & Goldfield, A. (1987). Easy to hear but hard to understand: A lip-reading advantage with intact auditory stimuli. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lip-reading* (pp. 97–113). London: Erlbaum.
- Rosenblum, L. D. (2005). Primacy of multimodal speech perception. In D. B. Pisoni & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 51–79). Malden, MA: Blackwell.
- Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, 26(2), 212–215.
- Summerfield, Q. (1979). Use of visual information for phonetic perception. *Phonetica*, 36, 314–331.

- Wang, Y., Behne, D. M., & Jiang, H. (2008). Linguistic experience and audio-visual perception of non-native fricatives. *The Journal of the Acoustical Society of America*, 124(3), 1716–1726.
- Wang, Y., Jongman, A., & Sereno, J. A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *The Journal of the Acoustical Society of America*, 113(2), 1033–1043.

Martinez-Paricio, V., Koreman, J., Husby, O., Eggesbø, J., & Bech, Ø. (2015). Consonant clusters in online L2 teaching: A multilingual approach. In J. Levis, R. Mohammed, M. Qian & Z. Zhou (Eds). *Proceedings of the 6th Pronunciation in Second Language Learning and Teaching Conference* (ISSN 2380-9566), Santa Barbara, CA (pp. 115-125). Ames, IA: Iowa State University.

CONSONANT CLUSTERS IN ONLINE L2 TEACHING: A MULTILINGUAL APPROACH

Violeta Martínez-Paricio, Jacques Koreman, Olaf Husby, Jardar Eggesbø Abrahamsen, Øyvind Bech, Norwegian University of Science and Technology

L1-L2*map* and CALST (Computer-Assisted Listening and Speaking Tutor) are two complementary platforms for second language pronunciation teaching developed at the Norwegian University of Science and Technology (NTNU) in Trondheim. L1-L2*map* is a multilingual database that allows comparison of the phoneme inventories of a large number of languages. CALST is a Norwegian computer-assisted pronunciation training system (CAPT), which uses the result from the contrastive analysis in L1-L2*map* to select different pronunciation exercises depending on the native language of the user. The online version of L1-L2*map* and CALST contains segmental information and exercises for practicing the perception and production of individual sounds and sound contrasts. But for speakers of many languages, consonant clusters in a foreign language are hard to perceive and pronounce, even if the speakers have no difficulty with the individual consonants. The present article presents the expansion of the multilingual platforms with a contrastive analysis of consonant clusters and exercises to train their pronunciation. The article also discusses some of the methodological challenges in incorporating consonant clusters in L1-L2*map* and CALST.

INTRODUCTION

In second language courses, it is common to encounter learners with very different language backgrounds, especially when the language course takes place in the country where the target language is spoken. For instance, learners with varying native languages usually participate together in the same Norwegian courses taught at Norwegian universities. We can find, for example, Vietnamese, Spanish, Italian, German, Polish, Thai, French, Russian and Mandarin Chinese speakers attending the same Norwegian course. In such a teaching context, it is often difficult to train students in the pronunciation of a second language (L2): the difficulties learners experience generally vary depending on their native language (L1). Furthermore, teachers do not normally have an in-depth knowledge of the phonetics of all the course participants' native languages and, therefore, they do not always know beforehand which are the typical pronunciation challenges for each speaker. Moreover, there is usually not enough time in class to give corrective feedback to individual language learners (Koreman *et al.* 2011). As a possible solution L1-L2*map* and CALST were developed, two online tools for L2 pronunciation training, which take into consideration the sound properties of the user's L1 and its differences with the L2.

THE ONLINE PLATFORM: CONSONANTS AND VOWELS

Languages often show extensive overlap in the sounds they use. But a target language often contains consonants and vowels which are not part of the native language phoneme inventory. These sounds may present problems to learners.

L1-L2map: Multilingual Segmental Information

L1-L2map (<http://calst.no/L1-L2map>) is a multilingual database that allows comparison of the phoneme inventories of any two languages (a potential L1 and a potential L2) from a set of around 500 languages. Most of the segmental data available in L1-L2map have been extracted from UPSID (Maddieson 1984) and LAPSyD (Maddieson *et al.* 2011), two databases where the phoneme inventories and some basic phonological information about several hundred languages can be consulted. In addition to this, phonemic information of additional languages which are of interest because of their linguistic properties or because they are spoken by large groups of immigrants in Norway has been incorporated in L1-L2map. This comparative database is implemented as a wiki, so that language experts can at any time include the phoneme inventory of a new language and/or correct the existing data. The main difference between L1-L2map and UPSID/LAPSyD is that only the former allows direct comparison of the phoneme inventories of two languages.

In L1-L2map this segmental information is presented in very basic phonetic charts, which are based on the International Phonetic Alphabet charts. A simple color-coding is used to display the sound similarities and differences between two languages. As shown in the following figure, the phonemes that are present in both the L1 and the L2 are indicated in green; the phonemes that only occur in the learner's L1 are coded in blue, and, finally, those that exclusively occur in the L2 are signaled in red (Figure 1).




Phonemes present only in the L1	
Phonemes shared by the L1 and the L2	
Phonemes present only in the L2	

Figure 1. Color coding in L1-L2map

The phonemes in red are generally important from the L2 teaching perspective, because speakers tend to experience greater problems in distinguishing and producing unfamiliar sounds (i.e., sounds which are absent in their native languages, cf. Lado 1957). Flege (1995) and others have demonstrated that not all unfamiliar sounds in L2 are difficult to learn. Particularly, sounds which do not have a similar counterpart in L1 often do not present great difficulties. We nevertheless believe that learners must be given the opportunity to familiarize themselves with all sounds which do not occur in the learner's native language, and have therefore also included

unfamiliar sounds which are not necessarily difficult to distinguish from other sounds in the target language. (For more arguments in support of using the contrastive analyses hypothesis in the elaboration of pronunciation teaching materials, see Husby *et al.* 2011; Avery & Eilrich 1992.)

To better illustrate how the comparison of phoneme inventories is carried out and displayed in our database, the chart in Figure 2 below presents the results of a contrastive analysis in L1-L2map comparing the phonemic inventory of Mandarin Chinese (L1) with Northern Peninsular Spanish (L2). From this chart, the teacher of Spanish can infer for instance that Mandarin Chinese learners of Spanish will probably experience difficulties with the perception and realization of the fricatives [θ j x], but not with [f s]. Similarly, Mandarin speakers can be expected to have difficulty with the phonemes /b d g/, also indicated in red in the chart. And in fact, difficulties with the realization of voiced stops by Mandarin speakers have been attested in the literature on Spanish L2 acquisition (e.g. Blecia & Esteve 2014 and references therein): Mandarin speakers tend to realize voiceless stops, which are closest to the target phonemes in their L1 (e.g. *estu[t]io* instead of *estu[ð]io* 'I study') (for a more detailed discussion on how the multilingual comparison of segmental information is carried out at L1-L2map, see Koreman 2013).

Languages

Sounds

Mandarin

Spanish

Username:

Consonants

Consonants (other)

Vowels

Language information

View positions

	Labial			Coronal				Dorsal		
	Bilabial	Labio-dental	Dental	Alveolar	Palato-alveolar	Retroflex	Palatal	Velar	Uvular	
Plosive	p	b		t	d			k	g	
	p ^h		t ^h	k ^h						
Nasal		m		n			ɲ	ŋ		
Trill				r						
Tap, Flap				ɾ						
Fricative		f	θ	s		ʂ ʐ	ç ʝ	x	χ	
Lateral fricative										
Approximant							j	w		
							ɥ			
Lateral approximant				l						
Lateral flap										

Figure 2. Lay-out of the result of a contrastive analysis comparing the phonemic inventory of Mandarin (L1) with Northern Peninsular Spanish (L2)

The phonemic level displayed in L1-L2map can be argued to capture insufficient pronunciation detail, since it does not reflect allophonic variation. For example, the chart above shows that Mandarin learners of Spanish are likely to have difficulty with the realization of the phonemes /b d g/, but it does not reflect the allophonic variation in the realization of these phonemes, which in

Spanish may surface as stops [b d g] or fricatives [β ð ɣ]/approximants [β̞ ð̞ ɣ̞] depending on the context. However, this is not a *big* problem from the teaching perspective: once the phonemes of a language are correctly identified, allophonic variants can be indirectly taught in exercises which contain words instantiating all possible allophones of a phoneme. For an accent-free pronunciation of the target language, allophonic variation must be given explicit attention, but at the level of communicative effectiveness, which is the main goal of our online platforms, this can be ignored.

Given the number of languages documented in our database, L1-L2*map* is a useful resource for language teachers who want to familiarize themselves with the phonemic inventory of the languages of their students. More importantly, by automatically establishing unfamiliar contrasts for different learners, the multilingual database becomes useful in yet another respect: its information can be automatically extracted and linked up to a computer-assisted pronunciation training system (CAPT). CALST, the *Computer-Assisted Listening and Speaking Tutor*, is such a system and uses the result from the contrastive analysis carried out in L1-L2*map* to offer different pronunciation exercises to the learners depending on their native languages. In particular, the phonemes of the L2 that are marked in red (unfamiliar to the user) will be linked to different types of sound contrasts exercises. Learners are directly presented with exercises and do not see the information in L1-L2*map*, which is intended for phonetically trained developers and language teachers.

CALST: Training New Sound Contrasts

CALST allows students to practice their pronunciation, listening and spelling skills. CALST is based on VILLE (Wik 2011), a Swedish CAPT system, and it is freely available online (<http://calst.no>). Learners can use it for free, not only during class hours, but whenever they have time to work on their Norwegian pronunciation skills. Importantly, CALST benefits from the contrastive analysis carried out in L1-L2*map*. The first time the learner logs in CALST, she must indicate her native and target language. Once this information is registered, the program automatically carries out a contrastive analysis and links the sounds in red in L1-L2*map* (i.e. the possibly problematic sounds) to specific exercises in CALST in which the sound is contrasted with other sounds with which it is likely to be confused. In this way, each learner gets to practice the unfamiliar sound contrasts in the L2 depending on her specific L1. At present, the listening and pronunciation exercises have been developed and recorded only for Norwegian. Because there is no accepted pronunciation standard in Norwegian, exercises have been developed for four main Norwegian dialects, with one male and one female role model for each dialect.

Figure 3 presents a screen view of one type of listening exercise (Listen & Click) in CALST. This exercise will be offered, for example, to a Spanish speaker (L1) who is learning Norwegian (L2). Recall from Figure 2 in the previous section that Spanish contains the underlying phoneme for the bilabial stop /b/ and, hence, we expect that Spanish speakers would not generally experience problems with the production/perception of such phoneme. However, Spanish phonemic inventory does not contain the labiodental fricative /v/. To train Spanish learners of Norwegian on the contrast /v/ vs. /b/, they are exposed to different minimal pairs where both sounds contrast in similar positions (word-initially or word-finally). The perception exercise illustrated in the screen view in Figure 3 is a so-called AXB exercise. In AXB exercises, the learner first hears the word on the left, with /v/, then the word on the right, with /b/, and after

that, she will hear one of the two words again and has to decide which of the two first words it corresponds to. Through this type of phonetic training with different minimal pairs, the learner can improve her ability in distinguishing these sounds (see Koreman et al. 2013 for a detailed explanation on other listening, pronunciation and spelling exercises at CALST).

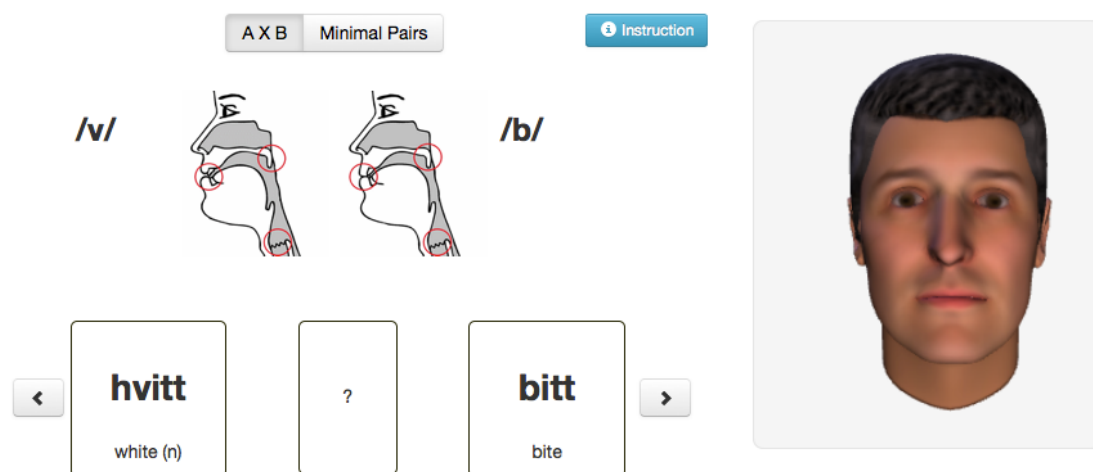


Figure 3. Screen view of a listening exercise contrasting a minimal pair beginning with /v/ vs. /b/

In addition to the exercises for new sound contrast, all users of CALST, independently of their mother tongue, have access to a list of general vocabulary exercises. In these exercises, learners can start to familiarize themselves with the sounds of the target dialect they want to acquire. These exercises allow practicing listening, pronunciation and spelling skills with a vocabulary consisting of around a thousand words and basic expressions.

Even if CALST currently contains only pronunciation exercises for several Norwegian dialects, it is important to highlight that it can easily be extended with other target languages (e.g. English, Spanish, French, etc.). Since the platform is modular, the extension of CALST is simple once exercise material is developed for other languages. CALST users who want to learn another L2 automatically take advantage of the multilingual contrastive analysis without any changes to L1-L2map.

CURRENT EXPANSIONS: CONSONANT CLUSTERS

It is not sufficient for L2 learners to practice the production/identification of individual sounds in contrast with similar sounds. The ability to combine consonants into clusters and/or to produce them in unfamiliar positions—different from the positions in which that same segment occurs in the L1—is a skill that needs to be practiced to ensure the intelligibility of L2 speakers.

Consider, for example, a Vietnamese learner (L1) of Norwegian (L2). Vietnamese and Norwegian both contain the phoneme /f/ in their inventory. However, /f/ displays a different distribution in the two languages: in English, the sound corresponding to the phoneme /f/ can occur in the onset and coda of a syllable (e.g. *[ff]eel*, *lea[ff]*); by contrast, in Vietnamese, this sound can only occur in an onset position (e.g. *[ff]ó* 'deal'). Since /f/ never occurs in the coda in Vietnamese, Norwegian words with [f] in the coda can be difficult to pronounce for a

Vietnamese speaker. Very often, instead of realizing a coda [f], speakers substitute it with an unreleased labial voiceless stop [p̚], especially at a beginner level (e.g. N. *sto*[p̚] instead of *sto*[ff], meaning 'stop' and 'fabric' respectively). As in the example, the substitution of the labiodental fricative [f] by an unreleased stop [p̚] in the coda can cause misunderstandings, particularly when both words exist in the target language, as it is the case for our example words. In sum, learners need to acquire not only the sounds of the L2, but also their particular distribution of each segment in the target language.

Another example of the importance of learning more than just individual L2 segments is attested in the interlanguage of speakers whose native languages allow only very simple (or relatively simple) syllable structures, while the L2 they want to master contains more complex syllables. These speakers often experience difficulties with the pronunciation of complex onsets and complex codas in the L2. For example, Spanish speakers (L1) are able to pronounce /s/ and /k/ when they occur in a simple onset position in English (L2) (e.g. */s/ea*, */k/ey*). Even the cluster itself is not a problem, as demonstrated by the Spanish word [es.ka.so] 'little, scarce' (with the dot indicating a syllable boundary). But when these segments form a complex onset as in */sk/ype*, Spanish learners of English tend to repair this complex onset, absent in their native language, by inserting a vowel before the two consonants. Instead of producing the monosyllable *skype* they often utter something closer to [es.kajp] or [əs.kajp], with two syllables. In a similar way, Thai speakers, whose native language bans such complex onsets, often avoid them by inserting a reduced vowel after the /s/, e.g. [sʰ.kaj] (Kenstowicz & Suchato 2006). This example shows in addition that, due to a native co-occurrence restriction in Thai, the final /p/ may be deleted after a glide in the interlanguage of these speakers (even though /p/ can occur syllable-finally in the language).

To summarize, native language restrictions seem to affect the production and perception of specific syllable types and consonant clusters in the L2 (Broselow & Kang 2013: 540). To enable training beyond single sounds, we decided to expand L1-L2map and CALST with multilingual information on possible consonant clusters and positional restrictions of segments. At the moment, our database contains this information for ten languages. In the remainder of this section we explain how this complex information can be now visualized in L1-L2map, we discuss the methodological problems we had to face when incorporating multilingual data on consonant clusters and explain how this information is used to develop new exercises in CALST.

L1-L2map: Consonant Clusters

L1-L2map offers a comparison of consonant clusters allowed in a pre-peak position (cf. onset) and a post-peak position (cf. coda) separately. Figure 4 below illustrates the comparison of consonant clusters as it appears in L1-L2map. In particular, this figure displays the comparison of possible pre-peak consonant sequences of two segments (C₁C₂-) in Norwegian (L1) and English (L2); that is, it lists and compares the possible consonant clusters of two segments that can occur before a sonority syllable peak in the two languages. The color coding we employ here is the same as in the segmental comparisons: (i) blue is used to mark complex C₁C₂- clusters allowed only in the native language of the learner (here: Norwegian), (ii) red indicates C₁C₂- clusters permitted only in the L2 (i.e. English) and (iii) green is used to signal C₁C₂- clusters which are common to both languages. The consonants listed vertically correspond to the first

consonant in the cluster (C_1); the consonants in the horizontal dimension indicate the consonants that can follow the first consonant (C_2).

The first row in Figure 4 shows that Norwegian and English both allow the clusters [pj-, pl-] before a vowel (green); it also shows that Norwegian in addition allows the cluster [pr-] (blue), while English permits the cluster [pɹ-], which does not occur in Norwegian (red). Ideally, this sequence and others highlighted in red (i.e. unfamiliar sequences for a particular learner, in this case a Norwegian learner of English) will be linked to specific production and perception exercises in CALST to support the learners in the acquisition of these new pre-peak consonant clusters. Once the English version of CALST is developed, a Norwegian learner of English, will, for instance, get exercises which would allow her to practice: (i) the clusters [pɹ-] as well as any other clusters that have the sound [ɹ-] as a second element in the cluster and (ii) *some* clusters containing the glides [j, w] in a second position (e.g. [kj-] and [kw-]). From the comparison in Figure 3, we can conclude that Norwegian and English are not generally very different with respect to possible C_1C_2 -, but making this automatic comparison allows Norwegian students of English to focus on those clusters which will probably be more difficult for them, rather than offering them training on all possible clusters.

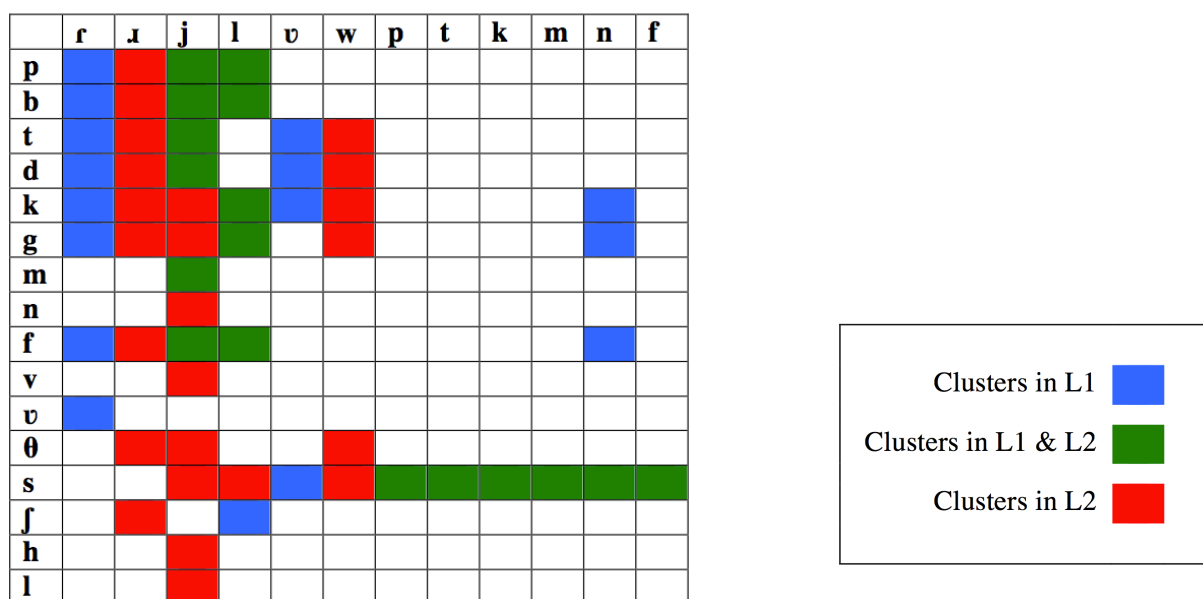


Figure 4. Screen view of the result of a contrastive analysis comparing pre-peak consonant clusters of two segments (CC-) in Norwegian (L1) with English (L2)

Introducing Multilingual Phonotactic Data in L1-L2map

This section offers some methodological considerations in the implementation of a contrastive analysis of consonant clusters. Since UPSID and LAPSyD do not offer complete lists of possible consonant clusters for each language, and we are not aware of other linguistic databases where this information can be easily consulted, we started recompiling specific syllabic information

from grammatical descriptions for the following ten languages: Norwegian, Spanish, English, Mandarin Chinese, Farsi, Vietnamese, Thai, Polish, Tagalog and Japanese. These languages were selected for both practical and typological reasons. On the one hand, we were interested in having data for languages representative of various language families with very different syllabic restrictions. On the other hand, we selected native languages which are quite frequent in Norwegian courses, so that the language information can already be of help for learners and teachers of Norwegian. We hope that *L1-L2map* will be expanded with additional data in the future in collaboration with other languages experts.

We first coded the segments and combination of segments that can occur *before* the syllable sonority peak in the above-mentioned languages (this is generally a vowel, but it can also be a consonant in languages which allow syllabic consonants) in the above-mentioned languages. Secondly, we coded the information related to specific segments and combinations of segments that can occur *after* the syllable sonority peak. We defined pre-peak and post-peak material (instead of onset and coda) to avoid analytical problems related to the particular association of prevocalic and postvocalic glides, which in some languages are analyzed as being part of the nucleus, but in others as being part of the onset for rising diphthongs or of the coda for falling diphthongs (see Smith 2007 for discussion). For teaching purposes, it is crucial to know whether both the L1 and the L2 admit, for example, a sequence like [Liquid+Glide] before the syllable peak. If linguists analyze the glide in L1 as being part of the onset, but in the L2 as part of the nucleus (together with the syllable peak), we are assuming that this is not relevant for teaching the pronunciation of the L2. The relevant fact is that the two languages are *equal* with respect to the presence of the sequence [Liquid+Glide]. When a speaker gets to practice the realization of complex sequences in the L2 in CALST, she will not get exercises focused on the sequence [Liquid+Glide]. Only speakers with an L1 that lacks the [Liquid+Glide] sequence at the beginning of a syllable would get such exercises.

By providing only the possible pre-peak and post-peak sequences of languages, we do miss some important linguistic generalizations since, very often, there are restrictions that hold over the whole rhyme (i.e. the nucleus and the coda) or over the onset and the nucleus. For instance, in Urban East Norwegian the vowels [e, ε] are in general blocked in front of the consonants /r/ and /ʀ/ and also before the glides /j/ and /v/ (Kristoffersen 2000:14). However, annotating all the possible syllables of a language –rather than just the possible pre-/post-peak sequences– would substantially increase the complexity of each language description, making the comparison between languages very difficult and redundant. Furthermore, we believe the L2 learner will indirectly learn this phonological, structural information when she is exposed to specific data and real examples in CALST. What is important for learners at the beginner level is to familiarize herself with the perception and production of unusual consonant clusters.

CALST: Developing Consonant Clusters Exercises

As explained previously, CALST currently contains exercises which allow learners to familiarized themselves with the pronunciation of general vocabulary, as well as with the discrimination and production of new sound contrasts not present in their native languages. Moreover, building on the information on consonant clusters, which has been implemented in *L1-L2map*, we are now expanding CALST has now been expanded to include pronunciation and listening exercises which allow learners to practice their production and perception of consonant

clusters too. We have now created the Norwegian material for exercises to correct five common repair strategies for consonant clusters: a) simplification (reduction of the cluster, e.g., [st-] > [s-], [t-] as in *top* or *sop* instead of *stop*), b) substitution (of one of the consonants, e.g., [br-] instead of [bl-]), c) metathesis (esp. in [st-] and [-ts] clusters, e.g. *tsop* instead of *stop*), d) prosthesis (vowel insertion *before* a cluster, e.g., *estop* instead of *stop*) and e) epenthesis (vowel insertion *in* the cluster, e.g. *setop* instead of *stop*). This material has been recorded and implemented in exercises in CALST.

A possible listening exercise (Listen & Click) for the Norwegian version of CALST is a minimal pair exercise. This exercise is similar to the sound contrast exercise displayed in Figure 3, except that the learner now only hears a single word. The learner is visually presented with two word alternatives, one of which contains an unfamiliar cluster (signaled in red in L1-L2map), while the other has undergone one of the repair strategies attested in L2 interlanguage (see above). For instance, imagine we want to train a particular user in the acquisition of the cluster [st-], which is banned in many non-Germanic languages. The learner will see on the computer screen the Norwegian words <sta> - <ta> (E. “stubborn”, “take”) or <sta> - <sa> (E. “stubborn”, “said”), while one of the words is pronounced by the virtual tutor. The learner has to choose the word she is hearing by clicking on the corresponding button on the computer screen.

Vowel insertion cannot be corrected in listening exercises such as the exercise described above, because there are very few relevant minimal pairs in Norwegian (e.g. *bro* ['bru:] ‘bridge’ vs. *bero* [bə'ru:] ‘to rest, to wait’). Pronunciation exercises will, however, offer an opportunity to learn to avoid these repair strategies. In pronunciation exercises (Listen & Speak), learners can train the pronunciation of words with unfamiliar onsets or codas. After listening to the artificial tutor’s pronunciation of a word with the complex onset [st-], for instance, the user records herself and then plays the realization of the word pronounced by the artificial tutor again, followed by the recording of her own voice. By comparing these, she can self-evaluate her pronunciation. In future, we aim to incorporate automatic speech recognition techniques in CALST, so that the artificial tutor will be able to detect pronunciation errors. Automatic speech recognition will not only help to detect vowel epenthesis repairs (e.g [est] or [set] instead of [st-]), but also the use of other repair strategies by the learner.

Finally, spelling exercises (Listen & Write) further strengthen the learner’s association of the perceived clusters with their orthographic representation. For reasons of space, we refer the reader to the CALST platform for examples of Listen & Write exercises, under “Vocabulary”.

FUTURE EXPANSIONS AND CONCLUSIONS

L1-L2map and CALST arose to train L2 learners in the acquisition of unfamiliar segments and new sound contrasts. We have explained in this paper how both platforms have been expanded to also incorporate information about consonant sequences. Still, the acquisition of an L2 sound system requires more than learning the new segments and their phonotactic restrictions in the L2. Learners must also acquire a new prosodic system, including (i) the position and phonetic cues of stress (in cases where the L2 has stress), (ii) the tonal patterns (in languages with lexical tones, but also in pitch-accent languages) and (iii) the intonational patterns of the language (Broselow & Kang 2013).

In the future, we hope to extend L1-L2*map* and CALST with information about the position of stress, making use of the StressType database (<http://st2.ullet.net/>), which contains information about stress in a large number of languages. Since there is no universal system which enables a contrastive analysis of lexical tones and intonational patterns, this will be implemented for Norwegian in CALST without reference to a contrastive analysis. Additionally, we will develop exercises to train the perception and production of Norwegian pitch accent and Norwegian intonation.

Finally, as we indicated in the introduction of this article, we hope to develop CALST with exercises for other languages, so that learners of other L2s can also benefit from this multilingual approach to language teaching.

ACKNOWLEDGMENTS

We gratefully acknowledge the financial support from VOX, the Norwegian Agency for Lifelong Learning, as well as the Ministry of Children, Equality and Social Inclusion, the Ministry of Local Government and Modernization and the Ministry of Labor and Social Affairs.

ABOUT THE AUTHORS

Violeta Martínez-Paricio (prosodic typology, violeta.martinez-paricio@ntnu.no), Jacques Koreman (phonetics and speech technology, jacques.koreman@ntnu.no), Olaf Husby (Norwegian as a Second Language, olaf.husby@ntnu.no), Jardar Eggesbø Abrahamsen (phonology, jardar.abrahamsen@ntnu.no) and Øyvind Bech (developer, oyvind.bech@gmail.com) collaborate in the cross-disciplinary CALST project (Computer-Assisted Listening and Speaking Tutor) in the Department of Language and Literature at the Norwegian University of Science and Technology (NTNU) in Trondheim, Norway. The project employs a contrastive analysis approach to language learning. The research group has developed and implemented pronunciation-training exercises for Norwegian, but their multi-lingual platform for language learning makes the same exercises immediately available for other languages when learning content for other languages is inserted. The platform intends to make pronunciation training available both in and outside the classroom.

REFERENCES

- Avery, P. & Ehrlich, S. (1992). *Teaching American English pronunciation*. Oxford: Oxford University Press.
- Blecua Falgueras, B. & Esteve Martín, M. T. (2014). Realizaciones de las oclusivas y aproximantes del español por parte de hablantes chinos. Oral presentation at *CIFE*, Valencia, Spain.
- Broselow, E. & Kang, Y. (2013). Phonology and Speech. In J. Herschensohn & M. Young-Scholten (eds). *The Cambridge Handbook of Second Language Acquisition* (pp. 529–544). Cambridge: Cambridge University Press.

- Flege, J. (1995). Second-language Speech Learning: Theory, Findings, and Problems. In W. Strange (ed.) *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (pp. 229-273). Timonium, MD: York Press.
- Husby, O., & Øvregård, Å., Wik, P., Bech, Ø., Albertsen, E., Nefzaoui, S., Skarpnes, E. & Koreman, J. (2011). Dealing with L1 background and L2 dialects in Norwegian CAPT. In *Proceedings of the workshop on Speech and Language Technology in Education*. Venice, Italy. [http://folk.ntnu.no/koreman/Publications/2011/L1-L2map_ICPhS2011.pdf]
- Kenstowicz, M. & Suchato, A. (2006). Issues in loanword adaptation: A case study from Thai. *Lingua*, 116, 921-949.
- Kristoffersen, G. (2000). *The phonology of Norwegian*. Oxford: Oxford University Press.
- Koreman, J., Bech, Ø., Husby, O. & Wik, P. (2011). L1-L2map: a tool for multi-lingual contrastive analysis. In *Proceedings of the 17th International Congress of Phonetic Sciences*. Hong Kong [http://folk.ntnu.no/koreman/Publications/2011/L1-L2map_ICPhS2011.pdf]
- Koreman, J., Wik, P., Husby, O. & Albertsen, E. (2013). Universal contrastive analysis as a learning principle in CAPT. In *Proceedings of the workshop on Speech and Language Technology in Education* (pp. 172-177). [http://folk.ntnu.no/koreman/Publications/2013/CALSTsoundcontrasts_SLaTE2013.pdf]
- Maddieson, I. (1984). *Patterns of Sounds*. Cambridge: Cambridge University Press
- Maddieson I., Flavier S., Marsico E. & Pellegrino F. (2011). LAPSyD: Lyon-Albuquerque Phonological Systems Databases, Version 1.0. [<http://www.lapsyd.ddl.ish-lyon.cnrs.fr/>]
- Smith, J. L. (2007). Representational complexity in syllable structure and its consequences for Gen and Con. In S. Blaho, P. Bye & M. Krämer (eds). *Freedom of Analysis?* (pp. 257–280). Berlin: Walter de Gruyter.
- Wik, P. (2011). *The virtual language teacher*. Ph.D. thesis, KTH School of Computer Science and Communication, Stockholm, Sweden.

McCrocklin, S. (2015). Automatic speech recognition: Making it work for your pronunciation class. In J. Levis, R. Mohammed, M. Qian & Z. Zhou (Eds). *Proceedings of the 6th Pronunciation in Second Language Learning and Teaching Conference* (ISSN 2380-9566), Santa Barbara, CA (pp. 126-133). Ames, IA: Iowa State University.

AUTOMATIC SPEECH RECOGNITION: MAKING IT WORK FOR YOUR PRONUNCIATION CLASS

[Shannon McCrocklin](#), University of Texas Rio Grande Valley

Automatic Speech Recognition (ASR), the technology behind language learning technology such as *Rosetta Stone* and *Burlington English*, is also available through many dictation programs freely accessible on the devices that students are likely to already own or have access to. Such ASR programs empower students to work on their pronunciation on their own, getting feedback based on the transcription provided by the program (McCrocklin, 2014). This paper introduces the potential benefits of using ASR in the classroom (such as fostering learner autonomy and supplementing course work), explores a few of the programs/technologies available (Siri, Google Voice Search, and Windows Speech Recognition), provides ideas for utilizing ASR as part of a pronunciation class (such as ideas for guiding student work), and addresses challenges that could potentially develop when using ASR programs (such as student frustration).

INTRODUCTION

Automatic Speech Recognition (ASR), the technology behind language learning technology such as *Rosetta Stone* and *Burlington English*, is also available through many dictation programs freely accessible on the devices that students are likely to already own or have access to. Such ASR programs empower students to work on their pronunciation on their own, getting feedback based on the transcription provided by the program. This paper introduces the potential benefits of using ASR in the classroom, explores a few of the programs/technologies available, provides ideas for utilizing ASR as part of a pronunciation class, and addresses challenges that could potentially develop when using ASR programs.

Background Information

Students learning a second language often recognize a need or desire to work on their pronunciation. Many students do not know how to work on their pronunciation outside of the language classroom, however. Unfortunately for these students, pronunciation is often treated as the “Cinderella” of language teaching (Kelly, 1969, p. 87), downgraded as a teaching goal and often pushed aside in favor of other skills (Isaacs, 2009; Lang, Wang, Shen, & Wang, 2012). As teachers, we want to help empower our students to practice and improve their pronunciation on their own, but few strategies and tools are readily available for students to use and we struggle to help our students become autonomous in their language learning.

One of the causes of student dependence on instructors in pronunciation learning is likely the classroom experience itself. Pronunciation instruction is often heavily led by the instructor. Many pronunciation classroom activities still rely on the teacher to model correct pronunciation and to monitor, evaluate, and give feedback on student production. Pronunciation teachers also

often rely on drills or controlled production activities, giving students little room for free expression or communicative practice. These types of pronunciation classes seem unlikely to help students practice outside of class because students are not encouraged to develop skills or strategies for monitoring or evaluating their own pronunciation and are given very little room for free experimentation with their pronunciation. Of the ten main teaching techniques introduced by Celce-Murcia, Brinton, and Goodwin (2010) for teaching pronunciation as part of the Communicative Approach (p. 9-10), only one, “recording of learner’s production” even mentions making use of the student in the evaluation. Admittedly, it is quite difficult for students to monitor their own production. In order to monitor and correct pronunciation accuracy, students must be able to hear when they make mistakes, which requires students to create aural discrimination categories appropriate to the L2, while research has indicated that sounds in an L2 are filtered through the phonological system of the first language (Beddor & Strange, 1982; Blankenship, 1991; Flege, Munro, & Fox, 1993).

Yet feedback is vital to the success of autonomous learning outside of the classroom (Sheerin, 1997). One technology that can help provide feedback is Automatic Speech Recognition (ASR), which allows students to experiment with the language in a safe, private setting. “[ASR] is an independent, machine-based process of decoding and transcribing oral speech. A typical ASR system receives acoustic input from the speaker through a microphone, analyzes it using some pattern, model or algorithm, and produces an output, usually in the form of a text” (Levis & Suvorov, 2014, p. 1). Many people have had experience with an ASR program, either through automated telephone lines, Siri on the iPhone, or speech dictation programs. ASR is also built into many language learning programs such as *Rosetta Stone* (2013), or *Burlington English* (2014). When used for pronunciation training, ASR is a tool that empowers students to practice at their own speed, getting feedback from the words recognized. Students that are trained to use ASR for their pronunciation practice have heightened self-efficacy, beliefs in their abilities to learn autonomously, and may engage in more autonomous learning behaviors (McCrocklin, 2014).

This paper explores the use of dictation technologies for pronunciation learning. While many students may enjoy learning from a dedicated language learning software, such as Rosetta Stone, such programs are often cost prohibitive and do not allow for flexibility in content. In order to make the pronunciation practice more accessible financially, this paper explores a few of the ASR technologies for pronunciation practice that are freely available (on various devices) for a wide range of target languages.

AUTOMATIC SPEECH RECOGNITION TECHNOLOGIES

Siri

Siri is described by Apple as an “intelligent personal assistant” available on the iPhone (Apple Support, 2014). It not only allows you to dictate messages, but also understands commands to complete actions. People can feel as if they are talking to Siri because the program can answer questions and will ask a question in return if the program does not understand a command or request. To begin using Siri on the iPhone, press and hold the home button (the physical button at the bottom of the screen). Once you hear two quick beeps, Siri is ready to listen to commands or questions. Siri can understand and speak the following languages: Cantonese, English, French (France, Canada), German, Italian, Japanese, Korean, Mandarin, Spanish (Mexico, Spain) (Apple

Support, 2014). Any language listed is available in any country in which you may be residing. To change languages for Siri, the user must enable a new language in the settings on the iPhone (Apple Support, 2014).

Benefit: Talking to Siri can feel like a conversation and therefore can feel more natural than dictating into other programs.

Drawback: Work with Siri may be more difficult to submit as homework, but students could dictate emails to the instructor through Siri.

Recognition Level: Due to its fairly high levels of recognition, Siri may work best for lower and intermediate learners.

Google Voice Search

Google's Voice Search is available on mobile devices as well as PCs and Macs in over 50 different languages (Chowdhry, 2014; Moon, 2014). Some of the languages available are: English, French, German, Italian, Japanese, Korean, Russian, Spanish, and Brazilian Portuguese (Google Support, 2014). One exciting new feature of the Google Voice Search application is the ability to update the settings to allow the program to detect and dictate in five different languages at a time, instead of changing the settings each time the user wanted to switch languages (Chowdhry, 2014). To utilize Google Voice Search without the application, simply go to www.google.com. To the right of the search bar, you will see an icon of a microphone. Click on the microphone icon and enable the microphone to get started with voice searches.

Benefit: Google voice search is particularly good with short utterances (one or two words) which could be useful for working with minimal pairs.

Drawback: As part of a class, work with Google Voice Search may be more difficult to submit as homework. Students would likely have to copy each of the search results individually into another document.

Recognition Level: Due to its fairly high levels of recognition, Google Voice Search may work best for lower and intermediate learners.

Windows Speech Recognition

Windows Speech Recognition (WSR) is an automatic speech recognition program that is already installed on PCs that use the Windows operating system (Microsoft, 2014). Once WSR is opened, it can be used to dictate into other programs, such as Microsoft Word. To get started with Windows Speech Recognition, click the start button and search for "windows speech recognition". Once completing a few start-up pages, the program will become available as a tab at the top of the screen (see Figure 1). Press the microphone icon to turn on the listening function. WSR is available in the following languages: Catalan, Chinese (China, Hong Kong, Taiwan), Danish, Dutch, English (Australian, Canadian, British, U.S., Indian), Finnish, French, German, Italian, Japanese, Korean, Norwegian, Polish, Portuguese (Brazil, Portugal), Russian, Spanish (Spain, Mexico), Swedish (Microsoft Developer Network, 2014). Upon purchase of a new computer, Windows (and therefore the speech recognition) is only able to operate in a single pre-installed language. To use another language, the user must download a language recognition pack for each language desired (to search for the language pack you are interested in, see <http://www.microsoft.com/en-us/download/default.aspx>).



Figure 3. Image of PC computer screen with Windows Speech Recognition tab

Benefit: Windows Speech Recognition makes homework submission easier as students can dictate into a Word document and submit the file.

Drawback: Windows Speech Recognition relies on context to help guess words uttered and, as such, has a much harder time with single words or minimal pair work. Words will be best understood if embedded into sentences for grammatical context.

Recognition: Windows Speech Recognition is fairly sensitive to deviations in speech patterns and therefore offers lower rates of recognition. Intermediate learners are likely to find the program challenging. The program is best suited for intermediate-advanced learners, although with voice trainings with the program (right click on the tab at the top of the screen while the program is running, click “Configuration”, and click “Improve voice recognition”) students could make the program more accommodating.

INCORPORATING ASR IN THE CLASSROOM

While there is some potential to use ASR for suprasegmentals (perhaps looking at two syllable noun and verb pairs, such as “to exPORT” versus “an EXport” with stress shifts), ASR is primarily useful for segmentals because students receive feedback on sounds from the spellings in the dictation. ASR work will also be particularly useful as a supplement and follow-up to classroom work time. Students should be introduced to differences in sounds and be given practice in properly producing sounds before working with ASR so that when students run into difficulties with words or sounds (i.e. the program mis-transcribing speech) the students can use information and tips from the class sessions to continue practicing and improving their pronunciation.

Work with ASR is easily incorporated through homework assignments when using Siri or Windows Speech Recognition. Students can dictate into a Word document, submitting the file through a course management system, or email, emailing homework directly to the instructor. Work with ASR may also be potentially used during class time if the teacher has a computer lab available. One issue to be aware of, however, is that the microphones in computers and smartphones are often sensitive to ambient sounds and may struggle to function properly if

background sounds, such as other students simultaneously working with ASR, are present. ASR is likely to work best in a lab with full or partial booths that will block some of the noise from travelling and high quality microphones, such as those designed for call center use.

When working with ASR, students understand relatively quickly and easily how the transcription provided from ASR can be used for feedback. The larger issue with working with ASR is that it can be frustrating to use, even for native and highly proficient speakers. In my own teaching experiences, students want to try the words and phrases until they get every single word transcribed perfectly, but sometimes this is simply not possible. It is important to help your students understand that the technology is not perfect and that they do not have to get every word and sentence transcribed perfectly. I tell my own students that the program is useful because it can help them identify areas to work on and provides feedback through the transcription. Then, I recommend to my students that they try saying a particular word up to three times, but if, after three tries, the program still has not recognized the word, they should move on. Some other tips that might be useful:

- If you are working with minimal pairs, ask students to focus only on the targeted sound. For example, if the target sound is /i/ in “beet” and they are able to get “be” they have accomplished the goal.
- Similarly, ask students to focus only on the targeted word(s) in a provided sentence.

Discussing the limitations of the program can help students set realistic goals for working with the program. Your role as a teacher using ASR in the classroom may also need to change as you help motivate students to push through the struggles and celebrate with them when they achieve successes.

Ideas for Practice Activities

To begin working with an ASR program, students simply need to have something to say. Teachers can provide guide sheets or materials for practice or students can find or develop their own materials. For teachers designing guide sheets, consider using ASR to follow-up on the sounds introduced in class. You can have lists of minimal pairs (beet-bit, green-grin, etc.) that students have to read through or sentences with words using the target sounds. It may be useful in these activities to highlight or bold the targeted words so that students can easily see how the ASR practice lines up with the practice done in class. The guide sheet could also ask questions that allow students to freely answer; for example, it could ask students to describe what they did over the weekend or to describe what they see in a picture. Moving from the minimal pairs which are a controlled activity to free responses will allow students to ease into practice with ASR while still focusing on their pronunciation accuracy on targeted sounds.

Students could also bring in their own practice with ASR. You could ask students to find a favorite poem or famous speech that they could read to the dictation program. Students could also prepare a presentation for your class (or another class) and practice the presentation with ASR. One of the great advantages of dictation ASR programs, is that any content can be brought in for practice and teachers and students are only limited by their creativity in direction and content of practice.

Challenges

In order for Automatic Speech Recognition practice to be successfully integrated into a course, teachers should be aware that there are challenges to using ASR. First, students may have trouble getting access to speech recognition technology. Students may not have access to a smartphone or they may have a Mac when you were hoping to use WSR on a PC. One way of approaching this is to be flexible, allowing some students to submit work through email with Siri and others to use WSR on PCs. Students could also be required to use one type of technology if the teacher ensures that all students can gain access to the same technology through school resources. For example, if the school has a PC computer lab or computers to check out, WSR is fairly easy to access for any student. Please note, however, that, if students must use a computer lab, they may feel uncomfortable speaking to the computer when there are other people working in the lab quietly on other projects. Finally, with a smartphone, PC, or Mac a student would be able to access Google Voice and may be a great way of making the technology available to all students (although Google Voice Search does not allow students to easily capture and demonstrate completed practice which may make homework submission more difficult).

Students are also likely to get frustrated with the technology for being too sensitive or not catching enough errors to provide helpful feedback. When using a more sensitive program such as WSR in my own classes, this frustration led students to doubt the program or even doubt themselves. Some students in my classes took their computers to their roommates or friends that were native speakers and asked them to dictate to the computer. When the computer failed to transcribe perfectly, my students became doubtful of the program's ability to help them practice their pronunciation. On the other hand, some of my students began to doubt themselves. Some indicated that they began to doubt their pronunciation abilities after using the program because the program was unable to transcribe their speech accurately. While it is useful to have a program show students where they are making pronunciation errors, the great amount of negative feedback can be overwhelming, particularly in the first practice with the program. I found my role as a teacher become much more about encouragement and acknowledgment that I was asking my students to complete a task which was frustrating. As students pushed through with repeated uses of the technology, however, they became more comfortable using the programs and saw the benefit of practicing with ASR, noting it helped them recognize their pronunciation problems and allowed them to practice repeatedly with items until they could get it right.

One final challenge is that students may cheat and there is not necessarily any way for teachers to know if students typed in their pronunciation work. While this is a possibility, it is fairly easy to tell when students type a dictation in because the submission of work is perfect, all words are identified perfectly and there are no errors, which is unlikely if students actually use an ASR program. In my own teaching experiences, students were more likely to type answers when they were confused about how to use the program or were frustrated by the low rate of recognition at the beginning of practice. Taking time in class to solve technology issues at the beginning of the semester and encouraging students to visit office hours if they continue to struggle with the program will be useful in helping students overcome issues related to the technology itself. Accepting and praising practice work that has transcription errors can also help students understand that the value of the practice lies more in doing the practice than in showing a perfect transcription at the end.

CONCLUSION

Automatic Speech Recognition can be a powerful tool for empowering pronunciation students to practice and get feedback on their pronunciation outside of class (McCrocklin, 2014). Although there may be challenges to re-envisioning and re-directing a dictation program for pronunciation purposes, the advantages of the program use make it worth trying. While this paper explored three main ASR technologies that cover a wide range of languages, there are likely many more reasonably priced or free options for the specific language of interest. Through exploring your options for integrating ASR into your classroom, you can provide students a tool that will allow them to work on their pronunciation outside of class, supplementing your in-class teaching of pronunciation and enabling students to work on their pronunciation autonomously.

ABOUT THE AUTHOR

Shannon McCrocklin is an assistant professor at the University of Texas-Rio Grande Valley. She completed her Ph.D. in Applied Linguistics and Technology at Iowa State University. She holds an M.A. in Teaching English as a Second Language from the University of Illinois at Urbana-Champaign where she developed an interest in pronunciation teaching and applied phonetics and phonology. Her research focuses on improving pronunciation training for students and CAPT (Computer-Assisted Pronunciation Teaching). She has presented at CALICO, NCTE, PSLLT, and AAAL. shannon.mccrocklin@utrgv.edu

REFERENCES

- Apple Support. (2014). About Siri. Retrieved from <http://support.apple.com/kb/ht4992>.
- Beddor, P.S. & Strange, W. (1982). Cross-language study of perception of the oral-nasal distinction. *Journal of the Acoustical Society of America*, 71, 1551-1561.
- Blankenship, B. (1991). Second language vowel perception. *Journal of the Acoustical Society of America*, 90, 2252-2252.
- Burlington English. (2014). Burlington English: Meeting the challenges of adult language acquisition. Retrieved from: <http://www.burlingtonenglish.com/about.aspx>.
- Celce-Murcia, M., Brinton, D., & Goodwin, J. (2010). *Teaching Pronunciation* (2nd ed). Cambridge: Cambridge University Press.
- Chowdhry, A. (2014). Google's Voice Search For Android Now Recognizes 5 Different Languages At Once. *Forbes*. Retrieved from <http://www.forbes.com/sites/amitchowdhry/2014/08/22/googles-voice-search-for-android-now-recognizes-5-different-languages-at-once/>.
- Flege, J.E., Munro, M.J., & Fox, R.A. (1993). Auditory and categorical affects on cross-language vowel perception. *Journal of the Acoustical Society of America*, 95, 3623-3641.
- Google Support. (2014). Ok Google and voice search. Retrieved from <https://support.google.com/websearch/answer/2940021?hl=en>

- Isaacs, T. (2009). Integrating form and meaning in L2 pronunciation instruction. *TESL Canada Journal*, 27 (1), 1-12.
- Kelly, L.G. (1969). *25 centuries of language teaching: an inquiry into the science, art, and development of language teaching methodology, 500 BC-1969*. Rowley, Massachusetts: Newbury House Publishers.
- Lang, Y., Wang, L., Shen, L., & Wang, Y. (2012). An integrated approach to the teaching and learning of zh. *Electronic Journal of Foreign Language Teaching* 9(2), 215-232.
- Levis, J. & Suvorov, R. (2014). Automated speech recognition. In C. Chapelle (Ed.) *The encyclopedia of applied linguistics*. Retrieved from <http://onlinelibrary.wiley.com/store/10.1002/9781405198431.wbeal0066/asset/wbeal0066.pdf?v=1&t=htq1z7hp&s=139a3d9f48261a7218270113d3833da39a187e74>.
- McCrocklin, S. (2014). *The potential of Automatic Speech Recognition for fostering pronunciation learners' autonomy* (Doctoral dissertation). Retrieved from <http://lib.dr.iastate.edu/cgi/viewcontent.cgi?article=4909&context=etd>.
- Microsoft. (2014). Microsoft accessibility: Products: Windows Speech Recognition. Retrieved from <http://www.microsoft.com/enable/products/windows8/default.aspx>.
- Microsoft Developer Network. (2014) Language support. Retrieved from [http://msdn.microsoft.com/en-us/library/hh378476\(v=office.14\).aspx](http://msdn.microsoft.com/en-us/library/hh378476(v=office.14).aspx).
- Moon, M. (2014) Google Voice Search can now handle multiple languages with ease. *Engadget*. Retrieved from <http://www.engadget.com/2014/08/15/google-voice-search-multi-language-default/>.
- Rosetta Stone. (2013). Cutting-edge technology: The software that started it all. Retrieved from <http://www.rosettastone.com/features#link1>.
- Sheerin, S. (1997). An exploration of the relationship between self-access and independent learning. In P. Benson and P. Voller (Eds.), *Autonomy and Independence in Language Learning* (pp. 54-65). London: Longman.

Qian, M. (2015). Extracting minimal pairs automatically with word frequency and phonetic environment controlled: Introducing a program written in PERL. In J. Levis, R. Mohammed, M. Qian & Z. Zhou (Eds). *Proceedings of the 6th Pronunciation in Second Language Learning and Teaching Conference* (ISSN 2380-9566), Santa Barbara, CA (pp. 134-142). Ames, IA: Iowa State University.

EXTRACTING MINIMAL PAIRS AUTOMATICALLY WITH WORD FREQUENCY AND PHONETIC ENVIRONMENT CONTROLLED: INTRODUCING A PROGRAM WRITTEN IN PERL

[Manman Qian](#), Iowa State University

A computer program that automates minimal-pair selection was developed using Perl. With variables such as L1 background, word frequency and syllable environment controlled, the system can identify, select, and extract minimal pairs automatically from the Illinois Speech and Language Engineering Dictionary. The minimal-pair selection follows Swan and Smith's (2001) phonology guide. This guide was chosen as the theoretical framework because it values learner-centeredness and recognizes that students speaking different mother tongues struggle with distinct pronunciation errors. With the program, different minimal pairs are respectively generated for English learners from 22 different native language groups. Minimal pairs can also be easily generated for additional learner groups by the program if error lists for their native languages similar to those in Swan and Smith (2001) are input. This paper describes the workings of the program and reviews the program's affordances and limitations in reference to its pedagogical and research implications. Directions for future development are also discussed.

ACCESSIBILITY ISSUES OF MINIMAL PAIRS

Minimal pairs enjoy a long and bittersweet history in pronunciation teaching. They were greatly embraced when first introduced to pronunciation teaching but gradually frowned upon following the burgeoning of communicative teaching due to their lack of context (Brown, 1995) and potential non-major role in real-world miscommunications (Brown, 1995; Levis & Cortes, 2008). However, despite this skepticism, minimal pairs never disappeared from pronunciation teaching materials and have continued in extensive use as training stimuli in research experiments (e.g. Lambacher et al., 2005; Wang & Munro, 2004).

In investigations and practices involving minimal pairs, what pairs to select is an important decision. Practitioners and researchers in general pay attention to the functional load (FL) of minimal pairs since sounds with higher FL have been found to decrease one's speech intelligibility more (Munro & Derwing, 2006). Materials developers also wish to let learner needs inform their choices so that words are not selected randomly but in a principled way reflecting and catering to individualized learner problems. In so doing, the suitability of the selected words for learners with specific proficiency levels and linguistic backgrounds can be improved. Nevertheless, from practical perspectives such a selection process can become laborious especially if approached manually — as is frequently the case.

Today has witnessed a growing number of language teaching websites with ready-made pronunciation teaching materials including freely accessible preselected minimal pairs. However, these preselected minimal pairs are either incomplete and need to be expanded or overwhelming and need to be filtered and organized before they can be used for classroom teaching or research purposes. For example, some websites (e.g. shiporsheep.com, homespeechhome.com, speechlanguagetherapy.com) provide incomplete minimal pairs without explaining why certain pairs are selected while others are not. These minimal pairs serve as a good start for material developers but usually offer only a limited range of words for further selections. Other websites such as [the Higgins List of Minimal Pairs](http://thehigginslistofminimalpairs.com) provides a comprehensive list of minimal pairs. On the other hand, the list displays all the existing minimal pairs in an unfiltered and somewhat overwhelming fashion. In addition, the minimal pairs on almost all these websites are chosen with little attention given to specific learner proficiency levels or pronunciation errors. The only website (that I discovered up until the time of writing) which takes some learner differences into consideration is englishclub.com where minimal pairs are categorically arranged and presented according to word frequencies. However, it is unclear what references are used to prompt the classification, so the categorization accuracy is somewhat dubious. Also, even this website does not control for any other variable such as individual learner errors in selecting minimal pairs. Therefore, even though these minimal pair materials freely accessible to us at present can indeed save people from the great trouble of digging out pairs from a dictionary, researchers and teachers often still need to spend time further developing the materials before applying them to real-world uses.

Development Goals: Program Functions

In light of the problems described above, the paper introduces a self-contained program that is intended to make minimal-pair selection more efficient as well as more reflective of individual learner needs. Specifically, the program was developed with the hope to have the following functions:

- a) Ability to correctly identify all the minimal pairs from a dictionary
- b) Ability to extract and output minimal pairs with word frequency controlled
- c) Ability to extract and output different minimal pairs suitable for learners with different phoneme-level pronunciation errors

PROGRAM DESCRIPTION

The program was written in Perl. After reading sound pairs from input text files, the program automatically identifies, selects, and outputs minimal pairs from a dictionary that are directly applicable to classroom and research use. The workings of the program can be divided into three primary stages, each focusing on solving one question (see Figure 1).

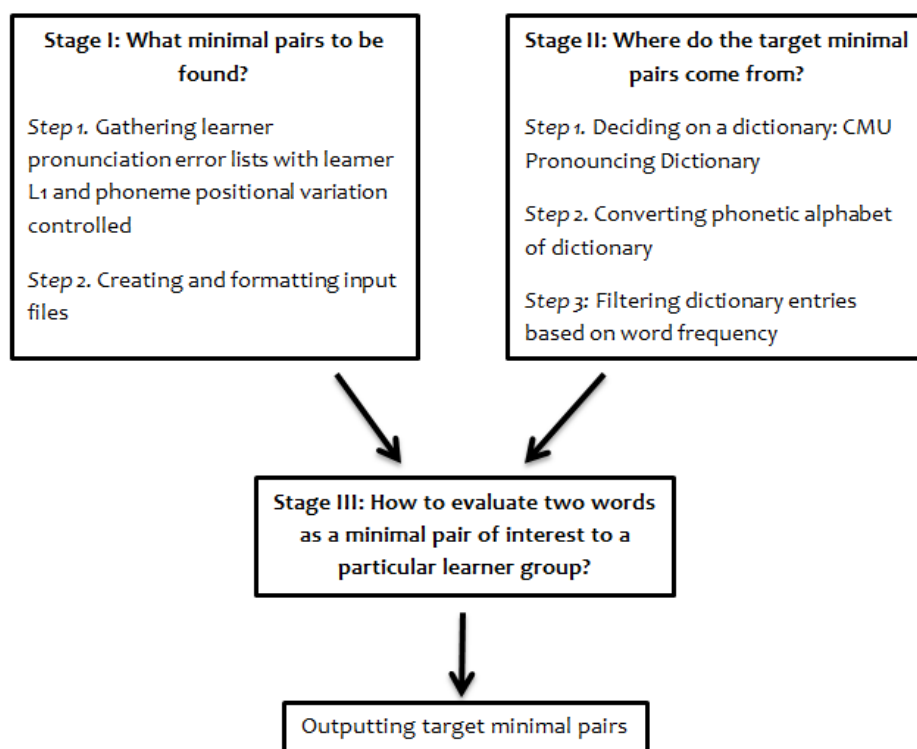


Figure 1. Workings of the system

Although the program was designed to automate minimal-pair selection, the activities involved in the first stage — preparing input text files — need to be performed manually, but this was the only stage where human intervention was required. To get the input files ready, two steps are followed: 1) finding out what sounds should be used to generate minimal pairs; 2) entering the target sounds into a text file following a specific format readable to the program. To decide on sound pairs that are pedagogically meaningful to build minimal pairs on, I followed Swan and Smith's (2001) phonology guide. This guide was chosen because it values learner-centeredness, recognizes that students speaking different mother tongues struggle with distinct pronunciation errors, and provides systematic and exhaustive error lists for learners from 22 different native linguistic backgrounds. By reading through the error lists, I respectively collected the sound pairs regarded to be challenging to each of the 22 learner groups. Next each set of sound pairs was entered into a text file following a format (see examples in Table 1) specifying the positional characteristics of the target sound pairs.

Table 1

Example Input Sound Pairs based on Swan and Smith (2001) and Anticipated Program Output

Reference in Swan and Smith (2001)	Input Sound Pairs	Anticipated Output
“/ʒ/ and /dʒ/ are rare in German. German speakers often realise them as /ʃ/ and /tʃ/ in English.” (p. 39)	dʒ-tʃ	dʒ-tʃ minimal pairs
“The lenis (‘voiced’) consonants /b/, /d/, /v/, /ð/, /z/, /ʒ/ and /dʒ/ do not occur at the ends of words in Dutch. Learners will replace them by their fortis (‘unvoiced’) counterparts: <i>Bop</i> for <i>Bob</i> ; <i>set</i> for <i>said</i> .” (p. 3)	b-p(initial)	/b/-/p/ minimal pairs whose initial phonemes are /b/ or /p/
“[For Japanese speakers] /t/, /d/, /s/ and /z/ often change before /ɪ/ and /i:/ as follows: /t/ become /tʃ/.” (p. 298)	t-tʃ (before[i:])	/t/-/tʃ/ minimal pairs with /t/-/tʃ/ sounds coming before the sound /i:/
“/ə/ in diphthongs such as /eə/, /ɪə/, /ʊə/ is usually replaced by the nearest Greek sound /a/.” (p. 130)	ə-a(after[ɪ])	/ə/-/a/ minimal pairs with /ə/-/a/ sounds coming after the sound /ɪ/
“Catalan, on the other hand, has a /z/-/s/ distinction similar to that of English, so there is no general problem. However, Catalan /z/ does not appear word-finally, so Catalans will say <i>face</i> for both <i>face</i> and <i>phase</i> , etc.” (p. 93)	s-z(final)	/s/-/z/ minimal pairs whose ending phonemes are /s/ or /z/

Stage 2 revolved around finding a source from which target minimal pairs would be extracted. The Illinois Speech and Language Engineering Dictionary (ISLEdict) attracted my attention as an ideal source for this project because the dictionary provides reliable pronunciation transcription for every word and is in the public domain. There are two reasons why the dictionary is believed to be reliable (“ISLEX,” n.d.): first, the pronunciation and lexical stress markings of 90% of its entries are from the Carnegie Mellon University (CMU) Pronouncing Dictionary, an authoritative reference that has been used for over 15 years; second, about 4,000 of the 137,000 entries in the ISLEdict have undergone manual checking and correction.

Nevertheless, the ISLEdict does not come ready for direct use for the purposes of this project, so the dictionary was processed by the program from two aspects in Stage 2. The first aspect concerns the phonetic symbol system used in the ISLEdict, namely Worldbet. Worldbet is a phonetic alphabet built with a primitive encoding system, ASCII, with the intention to cover and represent all languages in the world systematically. However, the phonetic symbols in the input text files established in Stage 1 were created on the basis of the British phonetic alphabet containing non-ASCII symbols, which is the phonetic alphabet system used in Swan and Smith (2001). As the mismatch between the two systems would certainly lead to inaccurate identification of minimal pairs, the phonetic symbols in the ISLEdict were first converted into the British phonetic alphabet version. The second aspect is word frequency. The ISLEdict contains 296,635 word entries, but many of the word entries are infrequently-occurring and may not be appropriate for English learners in general. Based on this assumption, the program was designed to account for word frequency when selecting minimal pairs by filtering out words of low usage frequencies. The New General Service List (NGSL), created by Charles Browne, Brent Culligan, and Joseph Phillips in 2013 to include carefully-selected high-frequency words in service to English as a foreign language in general (“NGSL,” n.d.), was adopted as the basis for the filtration.

Now the program knows what minimal pairs to seek for and where to go and find these minimal pairs, next in Stage 3 the program actually carries out the search by picking out all the minimal pairs that meet the requirements specified in the input text files produced in Stage 1 from the processed ISLEdict. In this process, two words were evaluated as minimal pairs if their phonetic spellings differ from each other in only one phoneme. These target minimal pairs were exacted and retained in third stage and finally delivered as output to users.

Using the program, different minimal pairs were respectively generated for English learners from 22 different native language groups (see Appendix A for some example output). Minimal pairs can also be easily generated for additional learner groups by the program if error lists for their native languages similar to those in Swan and Smith (2001) are input.

PROGRAM REVIEW

Affordances and Implications

The attractions of the program are twofold — it improves both the effectiveness and efficiency of minimal pair selection. From pedagogical perspectives, the program, in its controlling of three variables (i.e. learner L1 background, word frequency, and positional variation of the phoneme), facilitates conscious and effective minimal-pair selections that reflect diverse learner needs. This ultimately promotes learner motivation and outcomes (Nunan, 1988; Rodgers, 1969). In practice, the program is laborsaving as it completely automates minimal pair selection as long as one informs the program of what sound pairs to search for.

Additionally, the program also has good adaptability. In other words, the program is not limited to working on the basis of the NGSL as the word-frequency reference or Swan and Smith’s (2001) error lists as input. In effect, with no alternation necessary, the program can function well with any new word-frequency references and/or new error lists as long as the references or error lists are formatted similarly to the NGSL or Swan and Smith’s lists.

Given these features of the program, it has meaningful implications to researchers, materials developers, teachers, and students who take an interest in using minimal pairs. First, the minimal-pair output of this project is directly applicable to classroom and research use. Second, the output of the program can be specifically tailored to any particular learner or learner group if their phoneme-level pronunciation problems are known. One can also adapt the output by using a self-created corpus, for example a vocabulary list based on a specific textbook, as the word-frequency reference.

Limitations and Future Development

Despite the merits of the program, it has room for improvement and further development. First of all, although most of the program output is ready to use, the output comes with inaccuracies requiring manual removal. These inaccuracies are caused by two factors: 1) phonetic mistranscription in the ISLEdict and 2) the dictionary's provision of different phonetic spellings for one word entry associated with multiple speech varieties. As the phonetic transcription of entries in the dictionary is fairly reliable as previously mentioned, the first type of error is only occasional. More generally, misidentification of minimal pairs is attributed to the second factor. Appendix A lists some examples of the program output with erroneous items asterisked. Interestingly, all these asterisked word pairs belong to the second error type. For instance, among the 2,920 minimal pairs captured from the NGSL, all the pairs containing the word *about* are incorrect results. The reason is because *about* comes with two different phonetic spellings in the ISLEdict: /əbaot/ and /baot/. As it would be problematic to allow both forms to stay in the dictionary, for simplicity considerations, the program was designed in a way that only the last phonetic transcription was retained and compared with other words. This being said, it becomes apparent why the program paired *about* with words like *beat*, *bet*, *bit*, *bite*, *boat*, *boot*, and *but*. Similarly, as the word *a* comes with three phonetic spellings — /ə/, /ɑ:/, and /eɪ/, and only the last version was retained by the program, it evaluated *a-eye* and *a-owe* as minimal pairs. This was also the reason why *word*, transcribed both as /wərd/ and /wɜ:d/, was paired up with *would*. These several examples may arouse a question: would the problem be solved if the first phonetic transcriptions were picked in these situations? The answer is that, doing so would indeed solve the problem for these several examples but not for some other word entries in the dictionary because its listing of multiple phonetic transcriptions does not follow any systematic format — at least according to my observations for now, making it hard to solve the issue holistically by programming. Therefore, this drawback of the program does not appear to have an easier solution than manual analysis. However, experiments can be run to see whether removing the first or the last item from a queue of phonetic transcription forms would lead to a higher precision and recall rate.

Another important aspect about the program warranting further exploration is its usability. Currently the program can only be operated from the command box with Perl code, but this process is too technical, so further development should be directed to making the program easier to operate and access, from online for example. The usability of the program may also be enhanced if users are able to interact with the program and use their own selected or created dictionaries, word-frequency references, and learner error lists as sources.

Future uses of the program may be more productive if a newer version of Swan and Smith's phonological framework is applied to creating input error lists. For some problematic sound pairs

mentioned in the 2001 framework, few relevant minimal pairs were identified from the NGSL. Although more results could be produced if a reference list more encompassing than the NGSL were used, it could also be possible that errors collected from students over a decade ago no longer well reflect problems facing learners today. Nevertheless, the absence of minimal pairs corresponding to certain sounds may also be related to the fact that some falsely articulated words are simply not existent in dictionaries such as ‘*dat*’ (mispronunciation of ‘*that*’) and ‘*dere*’ (mispronunciation of ‘*there*’).

Lastly, the minimal pairs produced in the project are organized on the basis of the British phonetic alphabet only, while expanding this basis to be more inclusive of others, which is easy to implement from technical perspectives, can build up the applicability of the minimal pairs.

ACKNOWLEDGMENTS

I would like to thank Dr. John Levis and Dr. Evgeny Chukharev-Hudilainen for their guidance on my project development, my colleague Monica Richards for her recommendation of the ISLEdict as source for building the project, and the audience at my PSLLT presentation for their insightful feedback.

ABOUT THE AUTHOR

Manman (Mandy) Qian (mqian@iastate.edu) is a PhD candidate at Iowa State University. Her research interests include High phonetic variability input perception training, ASR-assisted pronunciation acquisition, and Applied computational linguistic analyses. She has taught a variety of classes at Iowa State, including Reading Skills for ESL Students, Listening Skills for ESL Students, Grammar and writing for ESL students, Advanced academic writing for ESL students, and English composition for native and nonnative students. mqian@iastate.edu

REFERENCES

- ISLEX. (n.d.). *International speech lexicon project*. Retrieved from <http://www.isle.illinois.edu/sst/data/dict/>
- Lambacher, S., Martens, W., Kakehi, K., Marasinghe, C., & Molhold, G. (2005). The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Applied Psycholinguistics*, 26, 227–247.
- Levis, J., & Cortes, V. (2008). Minimal pairs in spoken corpora: Implications for pronunciation assessment and teaching. In C. A. Chapelle, Y.-R. Chung, & J. Xu (Eds.), *Towards adaptive CALL: Natural language processing for diagnostic language assessment*, 197–208. Ames, IA: Iowa State University.
- Munro, M., & Derwing, T. (2006). The functional load principle in ESL pronunciation instruction: An exploratory study. *System*, 34, 520–531.
- NGSL. (n.d.). In *A new general service list (1.01)*. Retrieved from <http://www.newgeneralservicelist.org/>
- Nunan, D. (1988). *The learner-centred curriculum*. Cambridge: Cambridge University Press.

- Rogers, C. R. (1969). *Freedom to learn: A view of what education might become*. Columbus: Charles E. Merrill.
- Swan, M., & Smith, B. (2001). *Learner English: A teacher's guide to interference and other problems*. Cambridge: Cambridge University Press.
- Wang, X., & Munro, M. J. (2004). Computer-based training for learning English vowel contrasts. *System*, 32(4), 539-552.

APPENDIX A. EXAMPLE OUTPUT BY PROGRAM

*Note: * means erroneous output*

A total of 2920 minimal pairs were captured from the NGSL:

*1. a - eye	14. access - excess	2909. wipe - wise
*2. a - owe	15. accord - award	2910. wire - wise
*3. about - beat	16. account - amount	2911. wish - with
*4. about - bet	17. act - aunt	2912. with - worth
*5. about - bit	18. actor - after	2913. woman - wooden
*6. about - bite	19. ad - aid	2914. wood - word
*7. about - boat	20. ad - at	2915. word - work
*8. about - boot	21. ad - odd	2916. word - worth
*9. about - but	22. adapt – adopt	*2917. word - would
*10. about - doubt	2918. work - worth
*11. about - shout	2906. wing - wish	2919. yeah - you
12. abuse – accuse	2907. wing - with	2920. yes - yet
13. accept - except	2908. wipe - wire	

A total of 13 /u:/-/əʊ/ minimal pairs were captured from the NGSL:

1. blue - blow	4. new - know	7. news - nose
2. boot - boat	5. mood - mode	8. pool - poll
3. cool - coal	6. new - no	9. rule - role

- | | |
|-----------------|---------------------|
| 10. rule - roll | 12. through - throw |
| 11. shoe - show | 13. tune - tone |

A total of 7 /b/-/p/ minimal pairs beginning with /b/ or /p/ were captured from the NGSL:

- | | | |
|----------------|----------------|----------------|
| 1. back - pack | 4. beer - peer | 7. bowl - poll |
| 2. base - pace | 5. big - pig | |
| 3. bath - path | 6. bore - pour | |

A total of 25 /d/-/t/ minimal pairs ending with /d/ or /t/ were captured from the NGSL:

- | | | |
|----------------|----------------------|--------------------|
| 1. ad - at | 10. extend - extent | 19. seat - seed |
| 2. add - at | 11. grade - great | 20. side - sight |
| 3. and - aunt | 12. grand - grant | 21. side - site |
| 4. bed - bet | 13. hard - heart | 22. slide - slight |
| 5. bid - bit | 14. height - hide | 23. tend - tent |
| 6. cent - send | 15. inside - insight | 24. wed - wet |
| 7. cite - side | 16. odd - ought | 25. white - wide |
| 8. coat - code | 17. ride - right | |
| 9. dead - debt | 18. ride - write | |

A total of 3 /f/-/h/ minimal pairs with /f/-/h/ before /i:/ were captured from the NGSL:

- | | | |
|----------------|----------------|-------------|
| 1. fear – hear | 2. fear – here | 3. fee – he |
|----------------|----------------|-------------|

A total of 2 /n/-/m/ minimal pairs with /n/-/m/ after /i:/ were captured from the NGSL:

- | | |
|-----------------|--------------------|
| 1. scene – seem | 2. scream – screen |
|-----------------|--------------------|

Barrett, C. (2015). A sociocultural view of engagement in the music-based pronunciation classroom. In J. Levis, R. Mohammed, M. Qian & Z. Zhou (Eds). *Proceedings of the 6th Pronunciation in Second Language Learning and Teaching Conference* (ISSN 2380-9566), Santa Barbara, CA (pp. 143-153). Ames, IA: Iowa State University.

A SOCIOCULTURAL VIEW OF ENGAGEMENT IN THE MUSIC-BASED PRONUNCIATION CLASSROOM

[Catrice Barrett](#), University of Pennsylvania

Recent advancements in second language research point to the role of social interaction as a key factor in L2 development. In an era where cognition is no longer understood as a phenomenon confined to the inside of the head, Sociocultural Theory (SCT) offers insights about the ways in which language is regulated interpersonally for the purposes of mediating cognition, and hence L2 development. In this paper, I analyze students' collective engagement around a word stress rap activity occurring in a college intensive English course. The results show that students use linguistic resources to collaboratively resolve and attend to discrepancies in their abilities during the music-based activity. I conclude by discussing how these socially-mediated student behaviors translate into instructional opportunities to build learner self-sufficiency in the second language pronunciation classroom.

INTRODUCTION

Pronunciation instructors intend to shape students' second language development by setting "performance" or "behavioral" objectives in the lesson plan. This is what we are trained to do. Embedded in these objectives are our expectations about what students will *do*, how they will *respond* to our lessons. Despite our best intentions, this method of instructional design is undermined, at times, by the delicate and contingent nature of engagement, or the idea that what students *think* of the elements present in the interactional context of the lesson impacts what they *actually* do.

One potentially fruitful way to attend to the complexity of engagement is to look to recent scholarship, which has begun to underscore the influence of social interaction as an additional factor in the outcomes of the second language classroom (Atkinson, 2011). Sociocultural Theory (SCT) is predicated on the notion that "developmental processes take place through participation in cultural, linguistic and historically formed settings such as . . . peer group interaction" (Lantolf, 2000, p. 197). Through these processes, an individual's engagement is *mediated* not only by what happens inside the head, but also through the iterative co-construction of social and linguistic facts with others.

This paper examines interaction within a music-based word stress activity occurring during a pronunciation course in a college intensive English program. A series of classroom interactions are analyzed through the lens of SCT in order to put forth the argument that students use linguistic resources to mediate their engagement with music-based pronunciation in ways that result in drastically different roles and learning experiences for each individual.

Music and Second Language Pronunciation

At perhaps the most basic starting point, both music and language are universal to all human cultures. Both require and contribute to complex cognitive and motor processes (Vaneechoutte & Skoyles 1998). Additionally, duration, intensity (volume), pitch and tone are acoustic elements used in the production of both speech and music. The multifaceted connection between language and music has proven to be valuable in the ESL/EFL classroom. Our innate human affinity for music across linguistic borders has made it an accessible and inviting option for introducing learners to difficult or otherwise intimidating concepts (such as L2 pronunciation).

As it pertains to ESL pronunciation, numerous publications attest to the benefits of music-based pedagogy. Carolyn Graham's *Jazz Chants* (1978) popularized a musical approach to the teaching of communicative skills and prosody. Oftentimes, a chorus or other part of a song is repeated in such a way that reinforces language and aids in memorization (Mora, 2000). In an extensive teacher research dissertation, Terrell (2012) exploits the acoustic cues that are shared between language and music to teach pragmatic aspects of pronunciation to ELLs. These shared cues include: a high speech rate and intensity for anger; low voice intensity and rising pitch contour for fear; and slow speech rate and falling pitch contour for sadness. Additionally, Fischler (2009) noted intelligibility gains and increased metacognitive ability with respect to word stress placement among high school ESL students after they completed a series of rap activities from her *Stress Rulz* textbook. The results of these studies offer an added dimension of support for existing pronunciation research findings in favor of explicit prosodic instruction (Derwing and Rossiter, 2003).

Sociocultural Theory as a Lens for Student Engagement

Although, generally speaking, the literature on music-based approaches to second language instruction shows favorable results, these must be approached with caution as they are often aggregated and fail to reveal the nuance of individual student engagement. Moreover, the celebratory tone with which music pedagogies are often discussed could potentially obscure the struggles or divergent modes of engagement that lead students toward different learning paths, even as they seemingly pursue the same learning objective.

Sociocultural Theory (SCT) offers a set of principles that illuminate the relationship between differential modes of engagement in classroom interaction and associated cognitive processes. This is mainly because SCT posits that individual cognition is never simply relegated to the individual mind. Instead, cognition is primarily a *mediated* activity. That is to say that cultural and linguistic resources encountered in the social milieu function as tools that facilitate and iteratively re-shape the mental facilities, and from an L2 perspective, the linguistic development of the individual. Lantolf (2000), who has been a major champion of SCT in second language acquisition, maintains that *regulation* is one form of mediation. Regulation enables individuals to control resources (object regulation), their own behavior (self-regulation) and even the behavior of co-participants (other regulation) in order to achieve a goal. In a classroom context, the relevance of regulation quickly becomes apparent as individuals deploy language as a tool to make attempts to reach the learning objective collectively.

In this way, SCT presents an opportunity to capture the cognitive, affective and behavioral multidimensionality of second language classroom engagement (Fredricks and McColskey, 2012) in ways not possible through traditional, aggregated survey-based quantitative metrics. Tocalli-Beller and Swain (2007) emphasize this point. They analyze in-class student discourse and argue that adult ESL students “talk to learn” as they work on a task of interpreting riddles and puns. Likewise, Donato (1994) draws on SCT to demonstrate French learners’ joint construction of a reflexive verb form, a task that neither participant was able to complete alone. The scholarly trajectory paved by SCT research holds great potential for second language pronunciation pedagogy. As pronunciation instruction increasingly incorporates communicative approaches, more sociocultural research is needed in order to aid our understanding of how communicative engagement in pronunciation tasks and subsequent “development [are] based on collaboration . . .” (Vygotsky, 1987, p.210).

Research Questions

- How do students use language during the music-based pronunciation activity to regulate their own behavior and that of others?
- What do these forms of engagement reveal about differences in development among the students participating in the interaction?

METHODS

Participants and Setting

This paper documents the results of an action research study in which I took on the role of teacher-researcher in an intensive English program at a university in a large northeastern U.S. city. The setting of this study responds to Derwing and Munro’s (2005) observation of the need for “more classroom-relevant research” (p. 392) in the literature on second language pronunciation. The course, Intermediate Pronunciation, met twice a week for 90 minutes over a span of seven weeks. At least once a week, students engaged with a music-based activity as a means of practicing a target skill. These music-based interventions included choral repetition of poetry, rap and music lyrics, round robin language drills over an instrumental, creation of rap lyrics and cloze activities for listening practice. The class consisted of eight students from Arabic, French, Korean and Chinese L1 backgrounds.

Data Analysis

I conducted classroom discourse analysis on transcribed audio/video recordings of the music-based activities. This paper focuses interactions occurring during a particular word stress rap activity called “Change that Funktion” from Janelle Fischler’s *Stress Rulz* textbook (2006). The text is intended to help students remember the rule for the case when two words have the same spelling, yet word stress production changes based on the part of speech (or “funktion” – e.g. OBject [noun] vs obJECT [verb]). Such pairs of words are embedded into rap stanzas for students to recite over a beat. Nonverbal communication such as gaze, posture and gesture are also taken into account during analysis. I qualitatively evaluate occurrences of core discourse analytic constructs (such as timing and turn taking patterns) to inform my reading of how

students collaboratively engaged with the music-based activity to mediate their learning.

RESULTS

The student discourse occurring during the *Stress Rulz* activity reveals diverse forms of behavioral *regulation* or interactional “process[es] that students use to initiate and direct their efforts to acquire knowledge and skill” (Zimmerman, 1989, p.329). The musical nature of the activity lead to a participation structure that centered around two main speech events, which will be covered below: rehearsal time and performance time. Students rehearsed the rap text in pairs (with one group of three) as the teacher (myself) rotated and coached; later, students performed their rap for the class, round robin style, while still seated in pairs.

Regulating to Find New Support during the “Rehearsal Time” Speech Event

There were times at the front of the room when the real-time demands of the activity prompted students to retreat or reposition their roles. Some students’ post-activity self-reports confirmed that they faced challenges with the mechanics of the task. The speed of the beat and the instances when lines were overloaded with more syllables than the students could handle in time caused the most issues. This is not surprising, as the pace of rap is very uncompromising. It was precisely these challenges that revealed where students needed work with English prosody. In many cases their difficulties derived from not understanding how to shorten and reduce structure words and connect quickly across word boundaries. We later used activities to address these issues. However, to survive the moment of rehearsal time, students took some interactional liberties to gain a sense of self-efficacy with the rap task.

While seemingly simple on its surface, the task of repetitive listening and recitation of lyrics creates a cycle of engagement in which students constantly self-observe, self-judge and self-regulate (Zimmerman, 1989). Toward the goal of learner autonomy, such self-monitoring techniques are an encouraged component of contemporary pronunciation pedagogy (Miller, 2007). During the activity, some students could be observed regulating their own behavior through kinesthetic means such as tapping the desk, a foot or nodding their heads as they rehearsed. What the participation structure of pairwork added to this equation, however, is a layer that confirms the underpinnings of SCT; that is, that the language learning process occurs not merely inside the head, but it observably occupies the realm of social interaction as language mediates cognition. This happens as individuals’ language use reflects their ability to observe, judge and regulate themselves and others simultaneously.

One very salient example of interactional improvisations in the regulation process was observed through the scaffolding procedure of the activity. Initially, students were introduced to the *Stress Rulz* rap by way of the model rapper provided with the materials. After identifying the key stress features (contrasts) and taking the song home to rehearse privately, the model was eliminated and students were asked to rehearse to an instrumental with their partner in class. The removal of this scaffold exposed students who, for whatever reason (i.e. insufficient practice, etc.), had not achieved the skill to recite the lyrics fluently. At this point, socially-mediated regulation took hold. Whether they were willing or not, proficient students replaced the model that had been removed for the less proficient students.

Jung was a young businessman from Korea, whose quiet yet jovial energy contributed to the collegial interactions that occurred at the front of the room during rehearsal time. In his group of three, he seemed to be the most comfortable with reciting the rap lyrics. Mei, who was sitting next to Jung, clearly observed his skill and comfort with the lyrics, which differed drastically from her own. In response to what she noticed, she moved my audio recorder closer to Jung. Later, as Mei managed to recite a full line of the lyrics fluently, Jung reciprocated the appraisal of success by passing her the recorder (to which she laughed and pushed it away bashfully). Through this small playful form of object regulation, these two students are managing the interaction in a way that indicates their active observation and judgment of each other's progress with the task. Aside from indicating an awareness of being under the watchful eye/ear of a researcher, these pranks created a paradoxical space wherein play and laughter served to mitigate students' very palpable anxieties around the task of preparing to "perform."

Mei's estimation of Jung's skills contributed to her strategy of engagement in the interaction below. This was the second of three in-class rehearsals that the groups did without an explicit model (i.e. to the instrumental). Mei begins by confessing her trepidation to a nearby group member in Mandarin:

- 1 Mei: 我不知道是哪一句。
- 2 *I don't know which sentence we're supposed to do.*

- 3 Yixin: 跟上□奏。
- 4 (swirls hands in a circular motion) *Follow the rhythm.*

- 5 Mei: 我只会第一句。
- 6 *I can only do the first sentence.*

- 7 Yixin: 第一句?
- 8 (smiles) *The first sentence?*

The excerpt above begins with Mei regulating language by drawing on her L1, Mandarin. With this move, she has selectively enlisted the help of the only peer who shares her L1, Yixin. Simultaneously, she is engaging with affect by excluding Jung (a Korean speaker) from the participation framework. In other words, Mei's use of her L1 is significant in that it drastically reduces the number of listeners that can understand. Her inability to follow the task goes undetected thus saving face for her. Mei's comment also serves as a discursive indicator of low self-efficacy or "perceived capability to perform [an] activity" (Zimmerman, 2000, p.83). Self-efficacy, stemming from educational psychology, has traditionally been studied from a bounded, quantitative perspective. As can be seen here, however, a sociocultural approach to discourse analysis complements traditional approaches, as discourse also provides key insights into how Mei has covertly positioned herself as a struggling learner in this interaction. In the excerpt below, Mei seeks a suitable regulation strategy to overcome her struggles:

- 1 (♪♪ Music starts ♪♪)
- 2 (Jung raps alone and continues)
- 3 Jung: [To **proDUCE means**] we make a little more. But the [**PROduce**] is the
- 4 [lettuce] we buy at the store.
- 5 Mei: [To **proDUCE means**] (mumbles) [proDUCE]
- 6 [letter] (mumbles)
- 7 Jung: [To **inSULT**] express words as cold as ice. But the [**INsult**] is the comment that
- 8 isn't nice.]
- 9 Mei: [To **inSULT**] (mumbles) [inSULT is the comment that
- 10 isn't nice]

It is evident in many ways that Jung's abilities provide much needed support for Mei. While she cannot keep pace with the full length of the rap at the sentence level, she is able to chime in with the key vocabulary (contrast) words that we covered previously. For example, on line (5), Mei raps a few of the first words of the line "To proDUCE means." Quickly, she loses the pace, which Jung continues to sustain, and her intelligible utterances are exchanged for mumbles. Even the mumbling is worth mentioning because Mei mumbles in time with the beat. Essentially, Mei has developed a strategy of imitation. Lantolf (2000) explains that, from a SCT perspective, imitation is developmentally relevant in the sense that "it involves goal directed cognitive activity that can result in transformations of the original [cognitive] model" (p.203). Mei has observed both herself and Jung and judged the amount of linguistic load – including the amount of imitation- that is reasonable for her abilities, given the timing demands. As a result of this, she has committed to a set of priorities as part of her self-regulation process. The pattern of her utterances suggests that her order of priorities would look something like this: articulate focal stress words; keep pace with the beat; and least importantly - articulate every word precisely.

Interestingly, even though Mei's discourse shows a strong commitment to the focal stress words, there are still issues with her production. The purpose of the activity was for students to note and produce the stress contrast between verbs and noun pairs that are spelled alike. While Mei was able to do this successfully when pronouncing the words in isolation (prior to rapping), she consistently applies the *same* stress (not a contrast) to the pair of words during the rap. (proDUCE twice on line 5 instead of proDUCE/PROduce and inSULT twice on line 9). This corresponds to similar findings in second language pronunciation research in which scholars argue that correct pronunciation at a lower phonological level do not necessarily equate to success at a higher level (Celce-Murcia, Brinton & Goodwin, 2010).

Beyond discussing Mei's skill level, I am really seeking to emphasize the resourcefulness of her engagement process. Guided first by a plea for assistance in her L1 and later, by the more fluent discourse of Jung, she found a way to engage jointly with fellow students when the teacher removed support from the activity. In this way, she is able achieve a task that probably would have been impossible for her without the participation of others. Mei's engagement here shows that she is operating in her Zone of Proximal Development (ZPD), or the distance between her current level and the level at which she is able to function with assistance (Vygotsky, 1978).

Regulating to Shift the Participation Structure in the Performance Time Speech Event

After three rounds of rehearsal, the students were faced with the task of sharing or “performing” their raps for the rest of the class. The performances were brief and informal, as each student remained in their original seated position and was assigned only a couple of lines from the lyrics. Despite the casual format of the task, students were not eager to be singled out as the star rapper of the class. This was understandable and expected, especially given their varied levels of familiarity with and interest in the genre. In addition to using play as a mechanism for responding to the demands of the activity, students used their agency to re-arrange the participation structure of the speech event. For the purposes of performance, I had assigned individual sections of the lyrics for the students to recite individually. At the moment of performance time, however, the students’ preference for community and connection around the activity prevailed over my instructions:

- 1 Catrice: Now this is the instrumental. The guy [the model], he’s not singing. It’s just
- 2 us, okay?
- 3 (Yujia and Jung nod)
- 4 Catrice: You ready?
- 5 (silence)
- 6 (Jung scratches his head, chuckles and looks down)
- 7 (♪♪ Music starts ♪♪)
- 8 (Other students begin their turns)

- 9 Mei: (gestures to Jung and Yujia) Together! We together.
- 10 (Yujia looks toward Mei)
- 11 (Jung nods his head)

As with rehearsal time, I began the performance time speech event by reminding the students that we were no longer using a model. In retrospect, I realize that this may have only heightened their anxieties about performing, as evidenced by the silence on line 5 in response to my question “are you ready?” Despite what I read as initial trepidation by some students, many of them successfully recited their lyrics both in time with the beat and placing correct stress on the focal vocabulary words. Sahar and Yifei began, each reciting their lyrics separately. As Omar and Daniella picked up their respective turns, Mei decides to use language to regulate peer behavior. Essentially, she is reacting to her discomfort with the idea of performing solo. On line 9, as others are in the middle of their performance, she turns to her groupmates, in a somewhat panicked state and declares that they will perform together. Jung quickly consents nonverbally (line 11), and Yujia acknowledges the proposal with eye contact. Once the other students have finished their turns, Yujia, Mei and Jung jump right into the next section of lyrics, in unison:

- 1 Jung and Yujia: An **[OBject is something we can feel and see. To OBject expresses]**
- 2 a feeling that we don’t agree
- 3 Mei: **[OBject is something we can feel and see. To obJECT expresses]**
- 4 (mumbles)
- 5 (Yifei, Yujia and Jung laugh, shift in their seats)
- 6 (the group finishes the remaining lyrics)

7 Yifei: [(laughs and gestures to Sahar) We should have sang together.]
8

This group's deviation from the task instructions was evident, but so was their commitment to the task. Consistent with Wright's findings, the group "manipulate[d] the task process in order to make it manageable in their own terms" (as cited in Barkhuizen, 1998, p.87), which I certainly could appreciate. The decision to perform as a group was initiated by Mei, who, as noted previously, was a student who struggled the most with the activity. Although all of this would imply that Mei stands to benefit most from a group performance, something very unexpected happened. As line 1 shows, a slight divergence in stress production occurred as the group jointly recited the lyrics. This portion of the lyrics was supposed to highlight the stress contrast between *OB*ject as a noun and *ob*JECT as a verb. In actuality, Jung and Yujia misplaced stress on the verb, pronouncing it as *OB*ject. Despite her earlier struggles and her characteristic mumbles that follow, Mei was the only member of her group that correctly pronounced *ob*JECT, as shown on line 3. This outcome is certainly interesting as Mei's unassisted success may imply that in just a short amount of time, she is making strides toward moving to the next phase in her ZPD (i.e. being able to complete a task alone that was impossible for her previously). It is likely that the other students' blunder was attributable to the cognitive demand of keeping the quick pace of the beat while also needing to recall the stress rule. What is clear is that Jung and Yujia did not mispronounce the word due to a lack of knowledge. The regulation process of self/other observation and judgment was fully at work as several students (including Jung and Yujia) give a series of nonverbal cues on line 5 (laughter, etc.) to indicate that they noticed the mistake. Second language pronunciation research supports the idea that these moments of noticing one's own errors are a major prerequisite of development (Derwing & Munro, 2005; Smith & Beckmann, 2005).

DISCUSSION

Sociocultural Theory and other emerging social approaches to second language development (Atkinson, 2011) shed additional light on instances where a teacher's quick gaze might otherwise mislead him/her to believe that students in a group are participating fairly evenly. As the data show, even when there are discrepancies in student ability, some students are resourceful enough to regulate language in such a way that sustains their inclusion and progress toward the learning objective. To the extent that SCT illuminates differences in students' ability levels and resulting forms of engagement, it should equally inform pronunciation teaching and instructional design that move beyond a one-size-fits-all approach. A main lingering concern of this study, therefore, is the question of how to support and prioritize the individual needs of students.

Due to the fact that task observations in this study occurred during real-time, "survival mode" moments, student knowledge of their self-regulation strategies can only be assumed to be implicit, at best. Thus from a pedagogical perspective the insights of an SCT approach are only a first step, or a discovery phase of learner regulation strategies (such as prioritizing the production of prominence patterns over correct articulation of individual sounds). The major value of these insights is the potential they hold for iteratively shaping and reshaping the instructional materials (rubrics, reflection guides, etc.) provided to students. In order to optimize the value of these insights, they must combined with existing L2 pronunciation instructional models that turn implicit knowledge into actionable, explicit metacognitive

strategies (He, 2011). Toward this end, Sardegna and McGregor (2012) emphasize the need to scaffold this process by providing “opportunities for students to monitor their performance during pronunciation practice and reflect on their outcomes” (p.183) following practice. Leading students toward their own awareness of regulation skills is the link that empowers them to function effectively, even when socially-mediated support is not present.

Methodologically, the application of SCT through classroom discourse analysis stands to contribute insight to existing (typically quantitative) measures of student engagement. If aggregated, the results of student assessment in this course might suggest steady upward progress for the entire class. The granular level of transparency made possible by through the SCT framework complicates the false positive of steady group progress, thus enhancing our ability to see the needs of individual students. Moreover, the qualitative findings of this data analysis can be used either in isolation or in triangulation with additional methods (i.e. self-reported survey data, etc.) to yield a more valid and robust portrait of classroom engagement. As pronunciation literature increasingly points to the importance of student perceptions about their behaviors and abilities (Derwing and Rossiter, 2002), SCT offers a means of comparing these perceptions with what students actually do in a classroom setting.

Although this study offers a valuable glimpse of the complexity behind students’ differential engagement in the Stress Rulz rap it is admittedly limited in terms of what it can imply about the students’ second language development in a more enduring sense. Toward this end, a more extended series of observations are offered in later chapters of the dissertation from which this paper was derived (Barrett, forthcoming). Nonetheless, the interactions presented in this paper serve to raise our awareness about the delicate relationship between our learning objectives and the many paths our learners may take to achieve them.

ACKNOWLEDGMENTS

I begin by thanking Christ who has unfailingly provided me with an overflow of talent and strength to do all things, including this work. I am also incredibly grateful to Fengbiao Zhang for his translation skills as well as Betsy Rymes, Diane Larsen-Freeman, Yuko Butler and Shondel Nero for their continued guidance and encouragement throughout the course of my dissertation. Finally, it is my sincerest hope that the time spent with my students was valuable for their growth on many levels.

ABOUT THE AUTHOR

Catrice Barrett is a Ph.D. candidate in Educational Linguistics at the University of Pennsylvania. Her areas of specialization are TESOL and sociolinguistics. She has taught and developed English as a second/foreign language materials at Wuhan University in China, Temple University and various non-profits in Atlanta and Philadelphia. In her role as instructor in the Educational Linguistics division at the University of Pennsylvania’s Graduate School of Education, she has coordinated the TESOL Workshop, provided academic writing support and instruction for international students as well as training for pre and in-service teachers in second language writing pedagogy. Her research focuses on the digital and mass-mediated literacy practices of transnational, multilingual youth. The same commitment to 21st century literacies that inspired her research has propelled Catrice to explore technology, music, media and

performance in her work with ESL students in the classroom. She can be reached by email at catrice@gse.upenn.edu

REFERENCES

- Atkinson, D. (2011). *Alternative approaches to second language acquisition*. London: Routledge.
- Barkhuizen, G. P. (1998). Discovering learners' perceptions of ESL classroom teaching/learning activities in a South African context. *TESOL Quarterly*, 32(1) 85-108.
- Barrett, C. (forthcoming) *Finding our rhythm: Contextualizing second language development through music-based pedagogy*. (Unpublished Doctoral Dissertation). University of Pennsylvania, USA.
- Celce-Murcia, M., Brinton, D., & Goodwin, J. M. (2010). *Teaching pronunciation: A course book and reference guide*. New York: Cambridge University Press.
- Cole, M., John-Steiner, V., Scribner, S., & Souberman, E. (1978). LS Vygotsky, mind in society: The development of higher psychological processes. *Edited Trans.*. Cambridge, MA: Harvard University Press.
- Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly*, 39(3), 379-397.
- Derwing, T. M., & Rossiter, M. J. (2003). The effects of pronunciation instruction on the accuracy, fluency, and complexity of L2 accented speech. *Applied Language Learning*, 13(1), 1-17.
- Donato, R., & McCormick, D. (1994). A sociocultural perspective on language learning strategies: The role of mediation. *The Modern Language Journal*, 78(4), 453-464.
- Fischler, J., & Jensen, K. (2006). Stress rule: Rap music pronunciation activities for the ESL classroom: Book/Audio CD. *Vermont: Pro Lingua Associates*,
- Fischler, J. (2009). The rap on stress: Teaching stress patterns to English language learners through rap music. *MinneWITESOL Journal*, 26
- Fredricks, J. A., & McColskey, W. (2012). The measurement of student engagement: A comparative analysis of various methods and student self-report instruments. *Handbook of research on student engagement* (pp. 763-782) Springer.
- Graham, C. (1978). *Jazz chants: Rhythms of American English for students of English as a second language*. New York: Oxford University Press.
- He, L. (2011). Metacognition in EFL pronunciation learning among Chinese tertiary learners. *Applied Language Learning*, 21, 1-27.
- Lantolf, J. P. (2000). *Sociocultural theory and second language learning* Oxford University Press Oxford.
- Miller, S. (2007). *Targeting pronunciation* (2nd ed.). Independence, Kentucky: Cengage.

- Mora, C. F. (2000). Foreign language acquisition and melody singing. *ELT Journal*, 54(2), 146-152.
- Sardegna, V. G., & McGregor, A. (2013). Scaffolding students' self-regulated efforts for effective pronunciation practice. *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference. Vancouver, Canada: Simon Fraser University*,
- Smith, J., & Beckmann, B. (2005). Improving pronunciation through noticing-reformulation tasks. *PTLC (Phonetics Teaching and Learning Conference) Proceedings, University College London*, , 27(30.7) 2005.
- Terrell, S. H. (2012). Unpublished Dissertation. *Elements of music and speech: A methodology to incorporate the elements of music into teaching pronunciation to speakers of English as a second language*.
- Tocalli-Beller, A., & Swain, M. (2007). Riddles and puns in the ESL classroom: Adults talk to learn. *Conversational Interaction in Second Language Acquisition: Empirical Studies*, , 143-167.
- Vaneechoutte, M., & Skoyles, J. R. (1998). The memetic origin of language: Modern humans as musical primates. *Journal of Memetics-Evolutionary Models of Information Transmission*, 2(2), 129-168.
- Zimmerman, B. J. (1989). A social cognitive view of self-regulated academic learning. *Journal of Educational Psychology*, 81(3), 329.
- Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology*, 25(1), 82-91.

Brinton, D., & Chan, M. (2015). What's hot, what's not: Insights from pronunciation practitioners. In J. Levis, R. Mohammed, M. Qian & Z. Zhou (Eds). *Proceedings of the 6th Pronunciation in Second Language Learning and Teaching Conference* (ISSN 2380-9566), Santa Barbara, CA (pp. 154-165). Ames, IA: Iowa State University.

WHAT'S HOT, WHAT'S NOT? INSIGHTS FROM PRONUNCIATION PRACTITIONERS

[Donna M. Brinton](#), Educational Consultant, Beverly Hills, CA

[Marsha J. Chan](#), ESL Instructor, Mission College, Santa Clara, CA

An important discussion forum in today's global Applied Linguistics community is the electronic mailing list (e-list), which provides practitioners with an online discussion venue for the exchange of information and ideas. This paper summarizes discussions held on a moderated invitational e-list comprising an international community of pronunciation practitioners. The authors, both members of the e-list, share highlights from their research into those discussion topics (strands) and discussion sub-strands (threads) that generated the greatest amount of discussion over a one-year period.

INTRODUCTION

An important discussion forum in today's global Applied Linguistics community is the electronic mailing list (e-list), which provides practitioners with an online discussion venue for the exchange of information and ideas. In this paper, we summarize discussions held on a moderated e-list comprising an international community of pronunciation practitioners. Membership on this e-list, which numbers around 170, is invitational, and is based on participants' research and publications in the field.

In the e-list, participants have the options of: (1) generating a new discussion strand by asking a question or stating an opinion; (2) responding to other participants' points of view on a previously-initiated discussion strand; (3) sharing knowledge/resources on a topic in response to queries put out to the e-list; and (4) passively participating in the e-list by reading others' e-list postings. Other options on the e-list fall under the "housekeeping" category; these include such tasks as the nomination of new members and their introduction to e-list members as well as the dissemination of email and other contact information among members.

As members of the e-list, our goal in this study is to share highlights from those discussion topics (strands) and discussion sub-strands (threads) that generated the most discussion over a one-year period. This paper follows on the study by Brinton and Goodwin (2006), which used the same pronunciation specialist e-list to investigate pronunciation specialists' views on World English, intelligibility, and pronunciation standards. As Brinton and Goodwin note, the e-list discussion displays "the pronunciation specialists' desire to sort out misconceptions and determine teaching priorities" (p. 32). Given this desire, and especially given the recognized status in the field of the e-list's membership, the information communicated in the discussion postings is of great potential interest to the more general public of pronunciation practitioners.

Research Questions

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

1. Which topics are of current interest to international pronunciation specialists?
2. Of these, which topics elicited the greatest amount of response from the pronunciation specialists?

METHODS

We opted to analyze discussion strands and threads from the one-year period August 2013 to August 2014. From our initial sampling, we narrowed our analysis to the four topics that generated the greatest amount of discussion using the following selection criteria:

1. Topics with a minimum of 7 discussants; and
2. Topics with a minimum of 13 exchanges.

Applying these criteria, we identified four discussion strands. We then downloaded the discussants' comments into a separate document to facilitate our task in summarizing the main ideas and created a separate document to capture the references shared by the discussants on the four topics (see Appendix).

RESULTS

As Brinton and Goodwin (2006) note, "Listserv discussions have a life of their own. Although an initial posting can spawn numerous responses, discussion is not orderly, nor are discussants constrained to comment on or answer the original posting" (pp. 31-32). This characterization certainly holds true for the e-list discussion strands that we analyzed, some of which (e.g., Pronunciation and the Common European Framework of Reference) morphed into new discussion strands (e.g., The role of pronunciation in speaking test ratings). The non-linear nature of the e-list discussion presents some challenges to the researcher attempting to make sense of the different threads in each strand. On the other hand, the authentic nature of the discussion allows for disagreements among specialists to be voiced and for attempts at consensus reaching which make for fascinating reading.

Sample Discussion Strands

The following discussion strands provide a flavor for the types of issues discussed by the pronunciation specialists and represent the main topics discussed by the specialists during the time period in question.

Table 3

Sample Discussion Strands on the E-list, 2014

▪ Pronunciation as a motor skill	▪ IPA vs. “IPA-like” symbols
▪ Perfect vs. relative pitch	▪ Charts for English pronunciation
▪ The role of pronunciation in speaking test ratings	▪ Measuring L2 proficiency
▪ Syllabification for pronunciation	▪ Exemplary qualitative research
▪ Is second language (L2) fluency more important than pronunciation?	▪ What is a vowel really?
▪ Pronunciation and the Common European Framework of Reference	▪ Mondegreens

E-list Strands Generating the Most Discussion

Applying the above-stated criteria for selection, our analysis of the discussion strands was narrowed to four topics, as detailed in Table 2. The synthesis of each topic follows.

Table 2

E-list Topics Generating the Greatest Amount of Discussion and/or Controversy

<u>Topic</u>	<u># of Participants</u>	<u># of Exchanges</u>
Perfect vs. relative pitch	10	22
Pronunciation as a motor skill	11	35
The role of pronunciation in speaking test ratings	10	15
Syllabification for pronunciation	7	15

Topic 1: Pronunciation as a Motor Skill

The e-list posting which initiated this discussion strand posed the question of whether pronunciation is first and foremost a motor or a cognitive skill. For the purposes of the discussion, a motor skill was defined as involving a motor or muscular component that must be learned and voluntarily produced to proficiently produce the sounds of the language. A cognitive skill, on the other hand, was defined as involving the formation of concepts in order to categorize sounds according to the phonology of the language.

E-list participants were quite divided in their opinions on this topic, with the minority opinion represented by the assertion that pronunciation is primarily a motor skill (though the cognitive component is indispensable). Arguments proffered in support of this view included the following:

1. 146 head-and-neck muscles are directly involved in speech.
2. These muscles must be coordinated and fine-tuned to perform their acts.
3. Deliberate practice and multiple repetitions are required for automatization of articulation to occur (Ericsson, Krampe, & Tesch-Römer, 1993). Once automatic, it runs smoothly without conscious attention.
4. These motor processes are stored in procedural memory, which is stable throughout life. As such, they cannot be forgotten.
5. With automatization of the speech processes, the brain is free to add emotion and expression to the content of speech (without having to think about the mechanics of the performance).

The majority opinion for this strand was represented by the assertion that pronunciation is primarily a cognitive skill (though the motor skills are certainly involved). Arguments in support of this view included the following:

1. In order to learn pronunciation, concepts are formed. It is the brain that stores these new concepts and controls their execution.
2. Unfamiliar L2 sounds are processed by two different networks—the conscious or declarative and the unconscious or procedural.
3. Anderson’s ACT theory (ACT-R Research Group, 2002-2013) applies. Developing L2 pronunciation skills entails acquiring a form of expertise that becomes increasingly automatized.
4. Ultimately, pronunciation is a complex interplay of physical, perceptual, cognitive, and psycho-social factors.
5. Research on severe hearing loss in adults provides evidence that acquiring or maintaining a sound system is not primarily a motor skill (Lane & Webster, 1991).
6. Learning the motor skills involved in L2 pronunciation is a very conscious activity (i.e., one requiring brain resources). With time it becomes unconscious.
7. The nature of uptake to automated procedures is important to understand. What learners attend to consciously when learning a new sound is different (and separate) from what the subconscious automated networks are processing.

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: “In fact, I’ve been a little bemused that we have been having this discussion. Of course pronunciation (in particular, finding the right articulatory settings) becomes highly automatized with practice. But developing an L2 pronunciation (including the relevant articulatory settings) entails acquiring a form of expertise. As with many other types of expertise, the way in which we develop it has to be accounted for in cognitive terms.”

Topic 2: Perfect vs. Relative Pitch

The query initiating this second discussion strand asked if speakers of certain languages (e.g., Vietnamese, Mandarin) have a higher incidence of perfect pitch—with perfect pitch defined as the ability to hear any note of the scale out of context and identify which note it is. Relative pitch, conversely, can be defined as the ability to identify a given note on the scale by comparing it to a reference note and identifying the interval between the two notes.

This discussion strand generated much less controversy than the previous one, with discussion participants citing the following research findings in support of the assertion that speakers of tonal languages do indeed have a higher incidence of perfect pitch:

1. NSs of tone languages have a far higher incidence of perfect pitch than NSs of non-tone languages (Deutsch, 2006; Deutsch, Henthorn, & Dolson, 2004).
2. Having absolute pitch and being a first language (L1) speaker of a tonal language may make it more difficult to learn the prosody of a non-tonal language given the importance of relative pitch
3. L2 learners can develop “echoic memory”—allowing the shape of the pitch to echo in their head and thus facilitate the learning process.
4. Musicians can learn tonal contrasts more efficiently than non-musicians (Kraus & Chandrasekaran, 2010).
5. L1 Chinese speakers tend to speak English with an overall higher pitch than NS due to the lack of voiced obstruents in their L1.

Discussion participants also ventured opinions about the application of these findings to the second/foreign language classroom, as follows:

1. Absolute pitch is not the issue. Learners just need to hear pitch patterns through a different filter.
2. We have yet to account for learners who are musicians and yet have amazingly unmusical prosody.
3. A useful strategy to aid learners in reproducing pitch is to have them follow the shape with their hands.
4. Leaving a pause between the “listen” and “repeat” stages helps learners to hear an accurate internal echo.

Topic 3: The Role of Pronunciation in Speaking Test Ratings

The query initiating this third discussion strand explained that a graduate student wanted to do her thesis on the role that pronunciation plays in speaking test ratings and thus desired recommendations of rubrics to assess ESL speakers' mastery of different features of English pronunciation. Recommendations of systems for rating speaking generally, in an overall, holistic way, were also sought.

It was noted that some pronunciation textbooks contain diagnostic checklists or speech profile forms that are used for initial, diagnostic purposes, e.g., Prator and Robinett's (1985) *Manual of American English Pronunciation*, Grant's (2010) *Well Said*, and Henrichsen, Green, Nishitani, &

Bagley's (2009) *Pronunciation Matters*, with the remark that diagnosis is not necessarily the same as evaluation.

Several rating systems that include components related to speaking were discussed on the e-list. The Student Oral Language Observation Matrix (SOLOM) (Center for Applied Linguistics, n.d.) is a rating scale that teachers can use to assess their students' command of oral language on the basis of what they observe on a continual basis in a variety of situations. Pronunciation is one of five factors of this instrument designed for school children. The Common European Framework of Reference for Languages: Learning, Teaching, Assessment, (CEFR) is a guideline used to describe achievements of learners of foreign languages across Europe and, increasingly, in other countries (Council of Europe, 2011). CEFR, which is not targeted for any particular language, emphasizes *Can-do* statements, what a learner can do in particular interactions; however, the descriptors do not deal specifically with pronunciation. The Pronunciation Scoring Guide in Chan's (2009) *Phrase by Phrase: Pronunciation and Listening in American English* was also offered as a system for rating pronunciation holistically. These rating instruments consist of five or six levels. Table 3 compares the Level 4 descriptors of these instruments. For comparative purposes, Level 4 of the International Civil Aviation Organisation (ICAO) Language Proficiency Rating Scale, used to assess those employed in air-traffic communications, is also included. This rating scale consists of six skill areas and six proficiency levels.

Table 3

Level 4 Descriptors of four oral assessment rubrics

<i>Rubric</i>	<i>Description</i>
CEFR B2+ Level 4 of 6	<u>In informal discussion with friends</u> : Can keep up with an animated discussion between native speakers. Can express his/her ideas and opinions with precision, present and respond to complex lines of argument convincingly. <u>Formal Discussion & Meetings</u> : As above + Can identify accurately arguments supporting and opposing points of view. Can and respond to complex lines of argument convincingly.
SOLOM Level 4 of 5	<u>Fluency</u> : Speech in everyday conversation and classroom discussions generally fluent, with occasional lapses while the student searches for the correct manner of expression. <u>Pronunciation</u> : Always intelligible, although the listener is conscious of a definite accent and occasional inappropriate intonation patterns.
ICAO Language Proficiency Rating Scale Level 4 of 6	Operational – Passing level. Produces stretches of language at an appropriate tempo. There may be occasional loss of fluency on transition from rehearsed or formulaic speech to spontaneous interaction, but this does not prevent effective communication. Can make limited use of discourse markers or connectors. Fillers are not distracting.
Pronunciation Scoring Guide Level 4 of 6	<u>Functionally competent</u> . The speaker demonstrates functional competence in pronouncing English. Speech is generally intelligible, especially with a concentrated effort at listening. Demonstrates adequate pronunciation of words and phrases to convey global meaning. May occasionally delete or add sounds

	to words and display some hesitations. May contain some serious errors in phonemes, word stress, intonation, and sentence focus that occasionally obscure meaning.
--	--

Assessing pronunciation requires a consideration of the interplay of segmental, suprasegmental, and delivery elements. Speaking and pronunciation rubrics have descriptors that rely on subjective interpretation. In addition, the type of speaking task can affect the pronunciation of a learner/examinee. To date, no detailed version of pronunciation rubrics for any high-stakes testing situations exists. E-list participants seemed to agree that it would be helpful to have descriptors that reflect more closely the processes that learners actually have to employ rather than the inputs they can be expected to handle and/or the tasks that can be demanded of them.

Topic 4: Syllabification for Pronunciation

The query that launched the fourth discussion strand asked for colleagues' reactions to the idea of displaying spoken syllabification instead of written syllabification, along with the syllable-stress code, in a book for learners whose need for English is mainly oral. For example, instead of the typical dictionary word division de·vel·op·ment·al, print de·ve·lop·ment·al [4-3].

Written syllabification is based mostly on etymological or morphological principles and keeps meaning intact for dividing words in written form, especially when writing by hand. On the other hand, spoken syllabification is based mostly on phonemic principles and maximal onset, and it provides a phonological surface structure and a phonetic plan for actually speaking an utterance.

Table 4

Syllabification in English

<i>Written</i>	<i>Spoken</i>
learn·ing	lear·ning
de·vel·op·ment·al	de·ve·lop·men·tal
psy·chol·o·gist	psy·cho·lo·gist
stretch·ing	stre·tching

This query raised questions about the basis for the syllabification. One problematic point is the property of ambisyllabicity. Should *lemon* be divided lem·on or le·mon? The <m> in this word seems to belong to both syllables (lem·mon), e.g., le(m)·on. Ambisyllabicity occurs when a syllable with a lax vowel is followed by a weak syllable with schwa and there is only a single consonant letter linking them. In contrast, ambisyllabicity does not apply to the word *demon*, de·mon, in which the first syllable is a tense vowel.

Another problem involves orthographic representation and phonemic constraints. Should *mistaken* be divided: mis·tak·en or mi·sta·ken? Should *backup* be divided back·up or ba·ckup? In these examples, the second syllabification, where the syllable-final consonant is shifted to the beginning of the next syllable, causes unchecked lax vowels [mɪ] and [bæ], which are

theoretically not permitted in English phonology. Furthermore, a syllable beginning with the letters <ck> is not permitted in written English.

Discussants generally supported the concept of teaching spoken syllabification and endorsed an approach to pronunciation that takes account of the perceptual reality of Selkirk's (1982) principle of maximal onset, in which word-final consonants getting shifted when the following syllable begins with a vowel. Some participants shared the ways in which they presented and tested syllabification and stress in class. This strand concluded with a suggestion that the discussant use the syllabification that seems intuitively right.

DISCUSSION

Being part of a pronunciation e-list provides opportunities for researchers and practitioners to share interests, materials, methods, and other information. In the time period of this particular study, the topics that generated the most discussion from the greatest number of participants were (1) Perfect vs. relative pitch, (2) Pronunciation as a motor skill, (3) The role of pronunciation in speaking test ratings, and (4) Syllabification for pronunciation. Through these electronic discussions, participants have been able to learn about relevant research that can be applied to teaching, learning, and writing of materials and assessments. Moreover, such a forum enables us to compare, challenge, debate, change and/or confirm ideas.

While the results in this study were gleaned from discussions in a closed e-list, the general public can find many opportunities to discuss pronunciation in open electronic discussion and social media groups. Below are other such groups in which the authors participate:

1. TESOL SPLIS Speech Pronunciation & Listening Interest Section [www.tesol.org > Connect > TESOL Interest Sections > SPLIS](http://www.tesol.org/Connect/InterestSections/SPLIS): <http://www.tesol.org/connect/interest-sections/speech-pronunciation-and-listening>
2. CATESOL TOP-IG Teaching of Pronunciation Interest Group: <http://bit.ly/top-ig>
3. LinkedIn: www.linkedin.com
Groups, such as Accent Reduction Specialists, Communication and Accent Neutralization
4. Facebook: www.facebook.com Pages, such as IATEFL's Pronsig page: <https://www.facebook.com/pages/Pronsig/460534014066126>

ACKNOWLEDGMENTS

The authors wish to acknowledge the anonymous participants who contributed to the e-list discussion and whose contributions they have summarized.

ABOUT THE AUTHORS

Donna M. Brinton (dmbrinton@gmail.com) 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.

Marsha J. Chan (marsha@sunburstmedia.com) is on the faculty of the English as a Second Language Department at Mission College in Santa Clara, California and president of a small business, Sunburst Media for Language Learners. 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 upwards of 1500 free instructional videos at <http://www.youtube.com/PronunciationDoctor>.

REFERENCES

- ACT-R Research Group. (2002-2013). *ACT-R*. Retrieved from <http://act-r.psy.cmu.edu/>
- 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.
- Center for Applied Linguistics. (n.d.). *Student oral language observation matrix (SOLOM)*. Washington, DC: Author. Retrieved from <http://www.cal.org/twi/EvalToolkit/appendix/solom.pdf>
- Chan, M. (2009). *Phrase by phrase: Pronunciation and listening in American English* (2nd ed.). Sunnyvale, CA: Sunburst Media.
- Council of Europe. (2011). *Common European framework of reference for languages: Learning, teaching, assessment* (CEFR). Retrieved from http://www.coe.int/t/dg4/education/elp/elp-reg/cefr_EN.asp
- Deutsch, D. (2006). The enigma of absolute pitch. *Acoustics Today*, 2, 11-19. Retrieved from http://www.lifesci.susx.ac.uk/home/Chris_Darwin/PerMuSo/pdfs/Deutsch_AP.pdf
- Deutsch, D., Henthorn, T., & Dolson, M. (2004). Absolute pitch, speech, and tone language: Some experiments and a proposed framework. *Music Perception*, 21(3), 339-356. Retrieved from <http://www.auditory.org/mhonarc/2004/save/pdf00001.pdf>
- Ericsson, K.A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100(3), 363-406. Retrieved from [http://graphics8.nytimes.com/images/blogs/freakonomics/pdf/DeliberatePractice\(PsychologicalReview\).pdf](http://graphics8.nytimes.com/images/blogs/freakonomics/pdf/DeliberatePractice(PsychologicalReview).pdf)
- Grant, L. (2010). *Well said: Pronunciation for clear communication* (3rd ed.). Boston, MA: Heinle/Cengage.

- Henrichsen, L, Green B., Nishitani, A., & Bagley, C. L. (2009). *Pronunciation matters*. Retrieved from <http://pronunciationmatters.com/index.php>
- International Civil Aviation Organisation. (n.d.). *ICAO Language Proficiency Rating Scale*. Montreal, Quebec: Author. Retrieved from www.jarfcl.at/fileadmin/user_upload/ICAO.pdf
- Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience*, 11(8), 599-605. Retrieved from http://www.soc.northwestern.edu/brainvolts/documents/krauschandrasekeran_nrn10.pdf
- Lane, H., & Webster, J. W. (1991). Speech deterioration in postlingually deafened adults. *Journal of the Acoustical Society of America*, 89(2), 859.
- Prator, C. H., & Robinett, B. W. (1985). *Manual of American English pronunciation* (4th ed.). New York, NY: Holt, Rinehart, & Winston.
- Selkirk, E. O. (1982). *The syntax of words*. Cambridge, MA: MIT Press.

APPENDIX

Sources Provided by E-list Participants.

Pronunciation as a Motor Skill

- Elbers, L., & Wijnen, F. (1992). Effort, production skill, and language learning. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, and implication* (pp. 337-368). Timonium, MD: York Press.
- Ericsson, K.A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100(3), 363-406. Retrieved from [http://graphics8.nytimes.com/images/blogs/freakonomics/pdf/DeliberatePractice\(PsychologicalReview\).pdf](http://graphics8.nytimes.com/images/blogs/freakonomics/pdf/DeliberatePractice(PsychologicalReview).pdf)
- Fredembach, B., Hillairet de Boisferon, A., & Gentaz, E. (2009). Learning of arbitrary association between visual and auditory novel stimuli in adults: The “bond effect” of haptic exploration. *Plos One*, 4(3). e4844. doi:10.1371/journal.pone.0004844. Retrieved from <http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0004844>
- Kjellin, O. (1999). Accent addition: Prosody and perception facilitate second language learning. In O. Fujimura, B. D. Joseph, & B. Palek (Eds.), *Proceedings of LP '98* (Vol. 2, pp. 373-398). Prague, The Czech Republic: The Karolinum Press. Retrieved from <http://olle-kjellin.com/SpeechDoctor/ProcLP98.html>
- Kjellin, O. (1999). *Five cornerstones for second-language acquisition--The neurophysiological opportunist's way*. Unpublished manuscript. Retrieved from http://olle-kjellin.com/SpeechDoctor/pdf/Five_Cornerstones.pdf
- Lametti, D. R., Nasir, S. M., & Ostry, D. J. (2012). Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback. *Journal of Neuroscience*, 32(27), 9351-9358.

- Lane, H., & Webster, J. W. (1991). Speech deterioration in postlingually deafened adults. *Journal of the Acoustical Society of America*, 89(2), 859.
- Lang, L. (2014). *Overcoming challenges*. Retrieved from <http://www.carnegiehall.org/BlogPost.aspx?id=4295002049>
- MacDonald, E. N., Johnson, E. K., Forsythe, J., Plante, P., & Munhall, K. G. Children's development of self regulation in speech production. *Current Biology*, 22(2), 113-117.
- Messum, P. M. (n.d.). *Articles on child speech research and teaching English pronunciation*. Retrieved from <https://www.sites.google.com/site/pmessum/downloads>
- Ojemann, J. G., Buckner, R. L., Corbetta, M., & Raichle, M. E. (1997). Imaging studies of memory and attention. *Neurosurgery Clinics of North America*, 8(3), 307-319.
- Pahud, E. (2014). *How to solve a technical or musical problem: Carnegie Hall master class with Emmanuel Pahud*. Retrieved from http://www.youtube.com/watch?v=_21jqCDiPFE
- Paradis, M. (2009). *Declarative and procedural determinants of second languages*. Amsterdam, The Netherlands: John Benjamins.
- Raichle, M. E., Fiez, J. A., Videen, T. O., MacLeod, A. M., Pardo, J. V., Fox, P. T., & Petersen, S. E. (1994). Practice-related changes in human brain functional anatomy during nonmotor learning. *Cerebral Cortex*, 4(1), 8-26.

Perfect vs. Relative Pitch

- Bidelman, G. M., Hutka, S., & Moreno, S. (2013). Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: Evidence for bidirectionality between the domains of language and music. *Plos One*, 8(4). e60676. doi:10.1371/journal.pone.0060676. Retrieved from <http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0060676>
- Deutsch, D. (2006). The enigma of absolute pitch. *Acoustics Today*, 2, 11-19. Retrieved from http://www.lifesci.susx.ac.uk/home/Chris_Darwin/PerMuSo/pdfs/Deutsch_AP.pdf
- Deutsch, D., Henthorn, T., & Dolson, M. (2004). Absolute pitch, speech, and tone language: Some experiments and a proposed framework. *Music Perception*, 21(3), 339-356. Retrieved from <http://www.auditory.org/mhonarc/2004/save/pdf00001.pdf>
- Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience*, 11(8), 599-605. Retrieved from http://www.soc.northwestern.edu/brainvolts/documents/krauschandrasekeran_nrn10.pdf
- Ludke, K. M., Ferreira, F., & Overy, K. (2014). Singing can facilitate foreign language learning. *Memory and Cognition*, 42(1): 41-52. Retrieved from <http://link.springer.com/article/10.3758/s13421-013-0342-5>
- Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience*, 10(4), 420-422. Retrieved from <http://www.soc.northwestern.edu/brainvolts/documents/WongetalNatureNeurosci2007.pdf>

The Role of Pronunciation in Speaking Test Ratings

- Cauldwell, R. (2007). Defining fluency for air traffic control. *Speak Out!*, 37, 10-16. Retrieved from <http://www.speechinaction.org/wp-content/uploads/2012/03/Defining-Fluency.pdf>
- Chan, M. (2009). Holistic pronunciation scoring guide. In *Phrase by phrase: Pronunciation and listening in American English* (2nd ed.). Sunnyvale, CA: Sunburst Media. Retrieved from http://www.sunburstmedia.com/Pron_scoring_guide_MChan.pdf
- Cowie, R., Douglas-Cowie, E., & Wichmann, A. (2002). Prosodic characteristics of skilled reading: Fluency and expressiveness in 8-10- year-old readers. *Language and Speech*, 45(1), 1-36.
- Kang, O. (2013). Relative impact of pronunciation features on ratings of non-native speakers' oral proficiency. In J. Levis & K. LeVelle (Eds.). *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference*. Aug. 2012. (pp. 10-15). Ames, IA: Iowa State University. Retrieved from <http://jlevis.public.iastate.edu/pslltconference/4th%20Proceedings/Kang%20PSLLT%202012.pdf>

Syllabification for Pronunciation

- The Chambers Dictionary*. (2011). London, UK: Hodder Education. Retrieved from <http://www.chambers.co.uk/search.php>
- The Chicks' Own*. (n.d.). London: Amalgamated Press. Retrieved from http://ukcomics.wikia.com/wiki/The_Chicks%27_Own
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Selkirk, E. O. (1982). *The syntax of words*. Cambridge, MA: MIT Press.

Crabtree, J. (2015). Putting it all together: From pronunciation analysis to pronunciation pedagogy. In J. Levis, R. Mohammed, M. Qian & Z. Zhou (Eds). *Proceedings of the 6th Pronunciation in Second Language Learning and Teaching Conference* (ISSN 2380-9566), Santa Barbara, CA (pp. 166-173). Ames, IA: Iowa State University.

PUTTING IT ALL TOGETHER: FROM PRONUNCIATION ANALYSIS TO PRONUNCIATION PEDAGOGY?

[Janay Crabtree](#), University of Virginia

Research on pronunciation has found that intelligibility and comprehensibility can change without a shift in accent (Munro & Derwing, 1995), but do TESOL teacher trainees (TTs) internalize this research, and how do they put this knowledge into practice to devise effective speech analyses and plans for language learners?

Eleven TTs analyzed the same speech sample from the George Mason Speech Archives website (Weinberger, 2014), and then wrote hypothetical lesson plans based on their analysis. Before the analysis activity, all TTs were made aware of the research and terms covering intelligibility, comprehensibility, and accent as well as functional load (Munro & Derwing, 1995, 2006). While some of the TTs wrote focused plans to develop greater comprehensibility, comments and features to address pronunciation in the TTs' analyses suggest that TTs are not quite sure what features are tied to accent and what features to focus on to make learners more intelligible & comprehensible. Findings suggest that more rigorous training and discussion of comprehensibility and intelligibility as well as unpacking of these terms need to be conducted for TTs and integrated into training materials.

INTRODUCTION

In a general TESOL theory and method class, students who are teachers in training (TTs) must discuss teaching oral skills as part of their component of teaching the four skills for ESL. Many of the students have not had a phonetics or phonology course; however, many of them are aware of the IPA. All of them have been or are language learners of a foreign language, and many of the students who take the course are non-native speakers of English. The speaking section of the four skills is designated as a time in the course for students to put many of the resources they have become familiar with and put them all together—from understanding communicative competence to pronunciation to lesson planning.

In previous years, students in the TESOL certificate program merely read the chapters in the speaking section of Celce-Murcia's Teaching English as a Second or Foreign language test, but that did not seem to prepare the TTs adequately for the practice they later complete. This year, in their TESOL theory and methods course, the TTs studied two different chapters covering pronunciation, discussed two different pronunciation in-take sheets as a guide for a pronunciation analysis, and read articles to understand the terms *intelligibility*, *comprehensibility*, and *accentedness* as well as *functional load*. The students discussed segments as having either high functional load--impeding comprehensibility or low functional load--having less likelihood of impeding comprehensibility (as defined by Catford, 1988 and in Munro and Derwing, 2006). However, when students returned pronunciation analyses and plans, for the most part, these documents are a tome of epic proportions—proportions that would probably overwhelm a live learner and also waste a great deal of time working on elements that do not

impede comprehensibility. Based on comments from the members in the class, it becomes evident that the teachers in training (TTs) do not or cannot let go of native-speaking expectations for language learners or cannot focus on just a few elements to help further the comprehensibility of the language learner. Past discussions and comments by TTs in the general TESOL method and theory practice class concerning expectations for oral skills sparked this research because a trend among TTs in their comments led me to believe that they conflate the terms *comprehensibility* and *accent* which led to questions regarding what TTs do when they analyze speech samples and develop plans.

Research Questions

Will teaching the terms *comprehensibility*, *intelligibility*, *accentedness*, and *functional load* help students to create speech analyses that are focused so as to maximize time spent on pronunciation work to benefit the learners in gaining comprehensibility?

Will TTs conflate the terms *accentedness* and *comprehensibility*?

What types of lesson plans do TTs create in order to maximize time spent for comprehensible gains by the language learner?

METHODS

During the 2nd part of the theory and method course, the TTs began focusing on methods to teach the four skills (reading, writing, speaking, and listening) as well as grammar and vocabulary. In the speaking section, the TTs are given an assignment to analyze speech from an adult language learner and prepare a lesson plan based on that analysis. Prior to this assignment, TTs read the Pronunciation and Teaching oral skills chapters in the *Celce Murcia* by Janet Goodwin and Anne Lazarton and are given the readings from Munro and Derwing (1995) in which the researchers tease apart the terms *intelligibility*, *comprehensibility*, and *accentedness*. The TTs are also given brief readings on functional load (Munro and Derwing 2006) as well as the pronunciation profile from Linda Grant (2001) and the scale of comprehensibility from Issacs & Trofimovich (2012). Not only were the TTs given the terms, but a full day of class was utilized to discuss these terms and the scales used by Grant and Issacs & Trofimovich. TTs were also aware of the rubrics used in the TOEFL oral scoring, as these and other rubrics for assessment were analyzed in a prior assessment section of the course.

Participants & Speech Sample

Since the TTs are not yet in their practice (i.e. working with live language learners) which comes later in their certificate program and since I wanted to have the same speech sample for the simplicity of grading and discussion, the TTs were given an assignment to analyze a recorded speech sample from the George Mason Speech Archive site. A final reason for utilizing this particular speech sample was that it was from a language for which none of the learners had L1 or L2 knowledge. Seven of the TTs were native speakers of English, two were native speakers of Russian (one an adult learner and the other self-identified as an unbalanced bilingual with equal oral fluency but only English writing proficiency), one a native speaker of Mandarin Chinese, and one self-identified as being an unbalanced bilingual of English and Polish with

French as a foreign language. The TTs whose L1 was not English were advanced users of English. Of the seven TTs who were native speakers of English, three were learners of Spanish, one of Italian, one of Arabic, one of French, and one of Hindi. All had studied at least one foreign language, and many had studied more than one language.

Materials

All TTs were asked to listen to the same speaker, Vietnamese Speaker 5 in the George Mason Speech Archive site (Wienberger 2015). The TTs were given both written and oral directions for this assignment. The written directions in the TTs course collab site were as follows:

Look at the George Mason Speech Archive website. Listen to Vietnamese speaker 5 as if this were your student. Complete a speech analysis/profile and lesson plan to address what you say in the analysis. You will most likely want to use Linda Grant's intake sheet and the suggested guidelines for L2 comprehensibility developed by Isaacs and Trofimovich. Please upload these into this tab as a document, but also bring them to Tuesday's class for a short discussion.

TTs asked many questions about the assignment, such as if there was a length limit on the analyses and plan. This was left up to the TTs, with the explicit instruction that they were to treat the recording as if it came from a student they were working with and to focus on the items that impede comprehensibility the most; they were told they should tailor plans with that in mind. They were also urged to remind themselves of everything they had read in their section of pronunciation by reviewing it and to apply this to their analyses and plans.

The eleven (TTs) were given approximately a week to analyze the speech below (Vietnamese male speaker, age 21):

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 TTs then turned in the analysis for a grade and also brought them to class to discuss as a group.

The website also provides the recordings transcribed in the IPA and gives some basic biographical background data about the participants. This particular speaker was a 34 year-old male with a length of residence in an English-speaking country (USA) for two years. The site also provides some contrastive analysis of generalizations between the native languages and what the developer of the site terms general American English (GAE). The TTs were asked not to look at the phonetic transcription (Figure 1) or generalizations provided by the site, but the TTs may have done so, and this will be further discussed in limitations.

[bli: kɔ̃n stela: as hɜ tu bʌɪn
 dɪs tʰɪns wɪθ hɜ frɒm ðə stɔ:
 sisə? sɒū:ns ɒf frɛs snɔ bɪʃ
 faɪ tɪsə? slæbʊː ʌbə blu: ɔi:s ɛ
 meɪbɪ ɛ snæʃ fɔ xɒ bɪdɔ
 bapː vi ɔsɒv nɪt ɛ smɒl blætɪs
 snɛkː ɛnə bɪʃ tɔɪ frɒʃ fɔ ðə
 kɪs ʃɪ kæ s:kʌpː dɪs tɪn ɪntɔ
 ædː tʌ ɪedː bæse ɛn vi vɪʊ
 ʊɒ mɪt hɛ wɛnsdɛɪ stə tʌɪn
 steɪʃən]

Figure 1. Transcription of passage analyzed by TTS (Weinberger, 2014).

TTs uploaded their analyses and plans into the class collaborative site and then brought them to the following week's class for discussion. While some of the TTs may have viewed the site's analysis, many may not have as the TTs were seldom in agreement as to what features to begin with or what plan of action for the speaker might be most important. The TTs also could not agree on how comprehensible the speaker was both in discussion and in their analyses, with some of the TTs stating that the speaker was a beginner, and others stating that the recording was almost completely comprehensible, denoting an advanced learner. Almost all TTs agreed that vowels were not one of the main elements causing issues with comprehensibility for this speaker.

Data Analysis

TTs plans and analyses were then read, looking for similar trends among them. The analyses contained more than fifteen different characteristics in the speech analysis. Only those that were mentioned by more than one TT were placed in the data set, so any trend could be discerned. TTs then brought in their analyses and plans to share with the larger group. In discussing some of the analyses, TTs were asked to discuss why they had chosen to include items that might not necessarily impede the speaker's comprehensibility.

RESULTS

The speech analyses and plans ranged greatly in length and breadth from lengths of one page to nine pages. The TTs also disagreed whether supra-segmentals or segments caused the greatest difficulty in both intelligibility and comprehensibility. The TTs also disagreed as to which segments were of greatest importance impeding comprehensibility as can be seen in the summary graph (Figure 2) of the features most noted by the TTs.

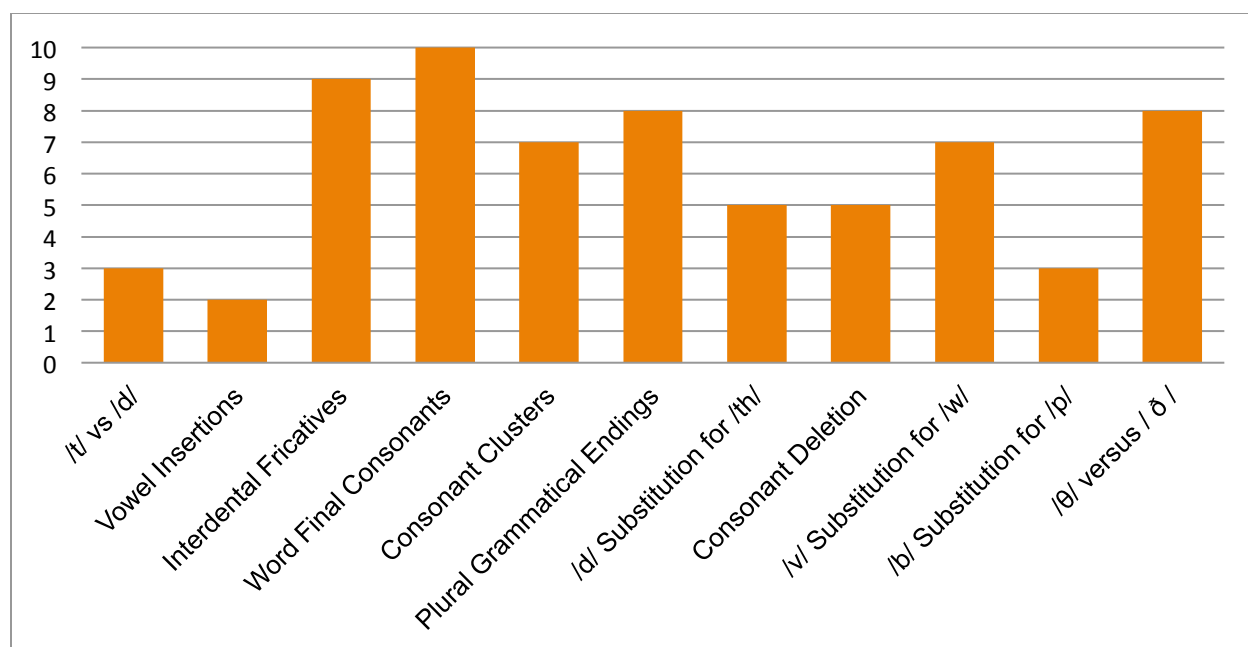


Figure 2. TTs speech analysis with most prevalent features presented as needing pronunciation focus.

Many of the TTs came up with lengthy analyses, but could not focus on just a few main items to develop a plan of action, as Linda Grant's pronunciation profile notes to do. For example, eight of the eleven TTs mentioned the speakers difficulty with /θ/ versus /ð/, but this distinction will probably not impede comprehensibility, and according to Catford has a low functional load (9-15). Of particular interest was that some of the participants noted that there were problems with interdental fricatives, but others made the distinction that there was a problem of voicing and voiceless or θ vs ð, as reflected in Figure 2. In fact, many of the participants noted both (interdental fricatives and θ vs ð) as being an issue and needing to be addressed. It is unclear on their assessments whether the TTs are not sure what interdental fricatives are or whether they see the replacement of the voiceless for the voiced as a different issue.

Another example of TTs pinpointing items that should not impede comprehensibility would be the /d/ substitution for /θ/ as well as a lack of grammatical endings. While these features may mark a speaker as non-native (i.e. being accented), it will probably not impede the overall comprehensibility of the speaker. Interestingly, some of the features that would be predictive of impeding comprehensibility were those least identified overall by the TTs, such as the vowel insertion or consonant deletion.

There were some elements in the analysis that most of the TTs noted that probably will have high functional load or impede comprehensibility, such as word-final consonants and consonant deletion, although some of the TTs did not further examine the environments in which consonant deletion was causing the greatest issue, and the TTs who noted word-final consonants as needing to be addressed did not always note if there was a pattern of deletion that needed to be particularly addressed and practiced such as final rhotic sounds in words such as *her*, *store*, *for* and *brother*. Another issue with this analysis is that, similar to the interdental fricatives and θ vs ð, TTs often note a problem with final consonants and consonant deletion, but it is unclear if TTs

are addressing the same issue. From their analyses, it is unclear that the TTs perceive this as the same problem or a different one; however, since they used two different terms, these two items (final consonant & consonant deletion) were placed in separate categories or as two different issues.

Based on these results, it became imperative to look at the data in a different way to see what percentage of the majority of the assessments were addressing items that might not push the learner to greater comprehensibility. Figure 3 depicts the percentage of all of the issues discussed out of all the features picked out by more than one TT. For example, if we add the items that are projected to have lower functional loads according to Catford (1988), then the TTs in training pick out over 30% of the features in the assessment as needing to be addressed that may not actually need to be addressed (θ vs δ , interdental fricatives, /d/ for /th/ substitution, grammatical endings, and /v/ substitution for /w/) meaning that TTs are neither maximizing their time nor their future students' time, if they continue with similar assessments and lesson plans developed to address features from those assessments.

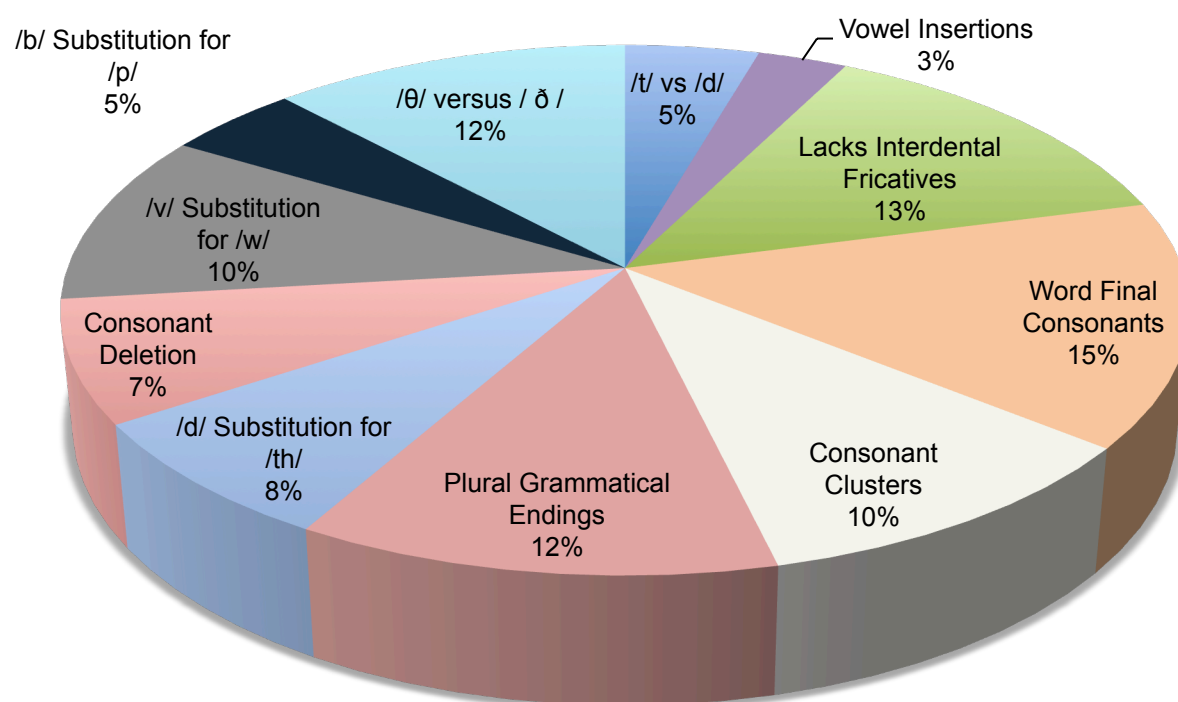


Figure 3: Percentage of all main features identified as needing to be worked on with the speaker according to TTs analyses.

DISCUSSION

From these results, the following items become clear: 1. TTs may still be using the native speaker as the pronunciation model, even subconsciously (as will be further discussed); 2. TTs

do not fully understand the difference between accent & comprehensibility, or again, are unwilling to let go of native speech as the default paradigm for their learners' pronunciation; 3. TTs need to more fully understand the term of functional load in order to efficiently develop lesson plans for learners.

In the post activity, TTs brought their assessments to class to share and discuss them. When I asked if the TTs were focusing on greatest comprehensibility and not necessarily accent, one TT stated that she thought that language learners will want to sound like native speakers of English, so she wanted to point out every feature that was not native-like. It must be noted that this TT is a native speaker of Russian and continued to explain that she felt that many of her students would be learning to conduct business and would be unfairly judged if they did not sound native-like. When I asked if that was a reasonable goal for all learners, many TTs stated that while not a reasonable goal, it might be the learner's goal.

What these results do indicate is that TTs do not fully understand the difference between accent & comprehensibility or their interplay even though these terms and their correlations were explicitly discussed in class, or the TTs are unwilling to let the native speaker as pronunciation model go. It will become important to tease these two apart in training teachers.

Also apparent is that the TTs are perceiving different features as having functional load or do not understand that term clearly. The Celce-Murcia chapters that the students read in their training focus on communicative competence, but never directly discuss clearly which features will cause the greatest impediment to that competence, although the Lazarton chapter does introduce terms such as *accuracy* and *fluency* while the Goodwin chapter creates awareness of the segmental versus the suprasegmental debate as well as advocating teachers to set realistic goals for the learners. Even having read these items, the TTs still seem to be pointing out features that are merely by-products of accent, not impositions to comprehensibility.

Over half of the TTs created lengthy lesson plans for minimal pair work with phonemes that may not impede the listener, a result that suggests the concept of functional load needs to be better covered in teacher-training materials to be able to help TTs apply it to their lesson plans, particularly regarding what causes the greatest lack of comprehensibility; otherwise, TTs run the risk of creating assessments and lesson plans that could potentially overwhelm the learners.

While the discussion of functional load may not induce the TTs to give up their conceptualization of the native-speaker as the goal for their learners pronunciation, as noted by the TT who stated that many language learners have a goal to sound native-like, it may clarify for the TTs more concretely just what realistic and sound goals are for the language learner so that they can put all of the information together to move from pronunciation pedagogy to sound practice.

Limitations

First, one of the limitations of this research is that the site provides both a phonetic transcription of the speaker and all speakers for that matter as well as generalizations about the speaker's accent for both language group and the specific individual. Some of the TTs may have looked at these generalizations to complete their analysis which is a limitation, in that this could have induced the TTs to create more analyses than they normally would have written for a live

learner. For future research, only the recording should be provided to ensure that the participants are not swayed by the site's analysis.

It must be noted that the teachers were working with a recording, so what they do with actual learners may be different, as the human element may force them to edit their analyses and plans. However, that one TT out of the eleven did edit her plan from her analyses to only three major elements leads me to believe that she envisioned the plan as being written for a real learner, as the instructions noted. If that is true as the assignment required, then it can be expected that the TTs will construct analyses and lesson plans similarly as they did in the exercise and overwhelm their learners.

ABOUT THE AUTHOR

When Janay Crabtree 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 and cycling. She is currently the lead faculty at the University of Virginia's Center for American English Language and Culture where she administrates the TESOL Certificate Program, teaches linguistics and ESL, and directs a summer English for Academic Purposes Program. jc9ne@eservices.virginia.edu

REFERENCES

- Catford, J.C. (1998). Functional load and diachronic phonology. In Y. Tobin (Ed.), *The Prague school and its legacy* (pp. 3-20). Amsterdam: John Benjamins.
- Grant, L. (2001). *Well said: Pronunciation for clear communication*. Boston, MA: Heinle & Heinle.
- Goodwin, J. (2001). Teaching pronunciation. In Marianne Celce-Murcia (Ed.), *Teaching English as a second or foreign language*, 3rd ed. (pp. 117-133). Boston: Heinle & Heinle.
- Issacs, T., & Trofimovich, P. (2012). Deconstructing comprehensibility. *Studies in Second Language Acquisition*, 34(3), 475-505. doi: 10.1017/S0272263112000150
- Lazarton, Anne. (2001). Teaching oral skills. In M. Celce-Murcia (Ed.), *Teaching English as a second or foreign language*, 3rd ed. (pp. 103-116). Boston: Heinle & Heinle.
- Munro, M., & Derwing, T., (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech*, 38, 289-306.
- Munro, M., & Derwing, T. (2006). The functional load principle in ESL pronunciation instruction: An exploratory study. *System*, 34(4), 520-531. doi: 10.1016/j.system.2006.09.004
- Weinberger (2014) . *Speech Accent Archive*. George Mason University. Retrieved from <http://accent.gmu.edu>. Accessed March 2014.

Deguchi, M. (2015). The role of pitch contours in teaching vowel length distinctions in Japanese. In J. Levis, R. Mohammed, M. Qian & Z. Zhou (Eds). *Proceedings of the 6th Pronunciation in Second Language Learning and Teaching Conference* (ISSN 2380-9566), Santa Barbara, CA (pp. 174-183). Ames, IA: Iowa State University.

THE ROLE OF PITCH CONTOURS IN TEACHING VOWEL LENGTH DISTINCTIONS IN JAPANESE

Masanori Deguchi, Western Washington University

Despite the significant difference in duration, vowel length distinctions have perpetually troubled learners of Japanese. Some studies have found that, while traditionally described as a “quantitative” distinction, the duration itself is not a reliable cue for discerning vowel lengths (e.g., Hirata, 2004). The present study therefore explores the possibility of capitalizing on a “qualitative” distinction, particularly different pitch contours, as a more reliable cue.

INTRODUCTION

Vowel Length

There are five vowels in Tokyo Japanese: /a/, /i/, /u/, /e/, and /o/. All of these vowels contrast in terms of their length, resulting in a total of ten distinct vowel phonemes (Shibatani, 1996). Therefore, words can contrast their meanings on the basis of vowel length alone as illustrated in the minimal pairs in Table 1. For example, *obasan* ‘aunt’ and *obaasan* ‘grandmother’ are different from each other only in term of the duration of a single vowel.

Table 1

Minimal Pairs Contrasting Vowel Length

Short Vowels		Long Vowels
/a/ vs. /a:/	oba <u>a</u> san ‘aunt’	oba <u>a</u> asan ‘grandmother’
/i/ vs. /i:/	oji <u>i</u> san ‘uncle’	oji <u>i</u> isan ‘grandfather’
/u/ vs. /u:/	su <u>u</u> ji ‘tendon’	su <u>u</u> ji ‘numerals’
/e/ vs. /e:/	se <u>e</u> kai ‘world’	se <u>e</u> kai ‘correct answer’
/o/ vs. /o:/	ho <u>o</u> soku ‘supplement’	ho <u>o</u> soku ‘natural law’

Since it is a phonemic property, the length distinction is extremely clear to native speakers of Japanese. In fact, phonetic experiments have found that Japanese long vowels are 2.4~3.2 times as long as their short vowel counterparts in their duration (Han, 1962; Ueyama, 2000).

Objectives of Present Study

Despite the substantial difference in duration, learners of Japanese, who do not have vowel length distinctions in their native language in particular, tend to have significant difficulties with such distinctions (Tajima, Kato, Rothwell, Akahane-Yamada & Munhall, 2008). Such learners generally have trouble in both production and perception (Oguma, 2000; Toda, 2003). In perception for example, learners tend to mistake long vowels for short vowels (Oguma, 2000).

While I argue that the results of the present study are applicable to production as well, I focus on discussing the perception of vowel length distinctions in the ensuing discussions. The main objectives of this paper are threefold. First, I identify and discuss the reasons for the above-mentioned difficulties; second, I propose a way to teach how to discern vowel length distinctions more effectively; third, I discuss the logic behind the proposed method of teaching.

DURATION VS. PITCH

Problems with Duration

It has often been argued that the primary cue for vowel length distinctions is duration (e.g., Fujisaki, Nakamura & Imoto, 1975). However, a fundamental problem arises from the fact that duration is a relative concept. In other words, there is no absolute long or short. A long vowel is long because it is longer than the short vowel counterpart. What this means is that it always requires comparison, specifically comparison between two tokens (i.e., minimal pairs). This is particularly problematic for Japanese since minimal pairs contrasting vowel lengths are quite limited in number (Vance, 2008). While it is true that pitch is also a relative concept (Ladefoged & Disner, 2012), comparison of pitch is done within a token, the pitch of a mora and the pitch of another mora within the same word (i.e., pitch contour). Therefore, minimal pairs are not called for when we capitalize on pitch.

To make vowel lengths even less reliable, long and short vowels in Japanese significantly overlap with each other in their duration at different speaking rates (Hirata, 2004). In fact, some phonetic experiments (e.g., Kinoshita, Behne & Arai, 2002) have found that native Japanese speakers rely on pitch contours in distinguishing vowel lengths when the durational cue is unreliable.

Based on the reasons that I just discussed, I argue that duration, at least duration alone, cannot provide a reliable cue for vowel length distinctions in Japanese. It is important to also note that, as with any contrastive/phonemic properties, the distribution of long and short vowels cannot be predicted by rules. What this means for Japanese instructors is that they cannot possibly teach their students in which phonological environments they expect to find long vowels to the exclusion of short vowels and vice versa.

Pitch as a More Reliable Cue

In this section, I demonstrate that, while duration does not seem reliable, there are rather systematic correlations between long vowels and the pitch contours associated with them. I further demonstrate in the next section that these characteristic pitch contours follow from general accentuation rules.

Japanese is said to be a pitch-accent language, where each mora is associated with either a high-pitch (H) or a low-pitch (L). The mora is the smallest prosodic unit, or a “beat.” In Japanese, a mora can be a single vowel (V), a consonant followed by a vowel (CV), the coda nasal (N), or, most importantly for our ensuing discussion, the lengthened part of a long vowel (R). In the word *kaiin* ‘members’ for example, there are four morae: *ka* (CV), *i* (V), *i* (R), and *n* (N); the first mora is pronounced with low-pitch, and the rest high-pitch: LHHH. It is crucial to note here that a long vowel consists of two morae. With this background in mind, let us now discuss vowel length distinctions in three different positions in words: word-initially, word-medially, and word-finally.

As shown in figures 1 and 2 respectively, a long vowel is associated with either a high-low (HL) pitch contour (a.k.a. “pitch drop”) or a low-high (LH) pitch contour (a.k.a. “pitch rise”) in the word-initial position. The pitch contours are indicated by blue dotted lines in the spectrographs.

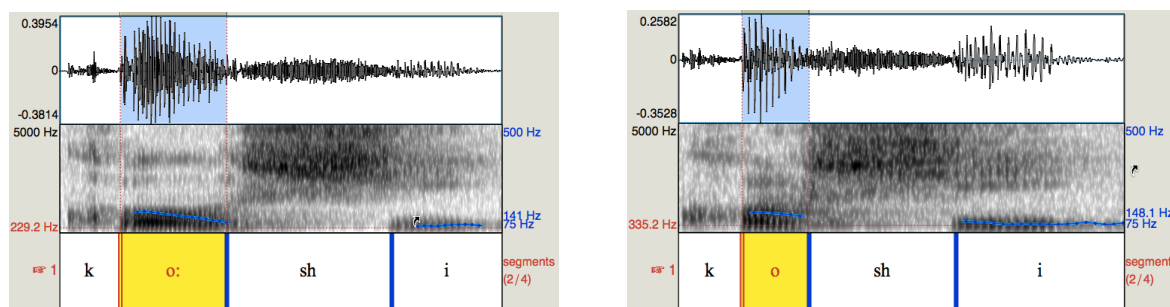


Figure 1. Length distinction word-initial position 1

Observe that, in the word *kooshi* ‘lecturer’ in Figure 1 (left), the pitch drops within the long vowel [o:]. Within this long vowel, the pitch dropped on an average of 44.33 Hz. Each token was recorded three times (see the appendix for the measurements). Recall that, since a long vowel is made up of two morae, a single long vowel can be associated with two distinct pitches. In *kooshi* ‘lecturer,’ the long vowel is associated with a high-pitch (average of 158.33 Hz) followed by a low-pitch (average of 114 Hz), resulting in a pitch drop (i.e., HL). On the other hand, no significant pitch change was observed in the short vowel in *koshi* ‘old paper’ in Figure 1 (right); the average pitch change within this short vowel was 11.33 Hz.

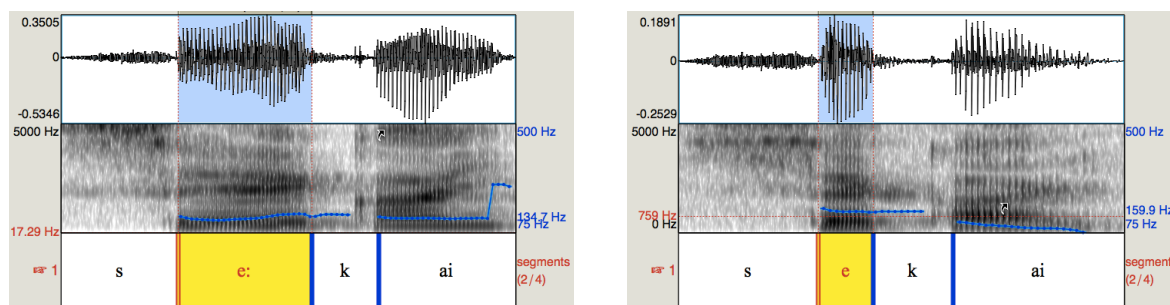


Figure 2. Length distinction word-initial position 2

Similarly, the pitch changes within the long vowel in the word *seekai* ‘correct answer’ in Figure 2 (left) as well; however, it changes from low to high, creating a pitch rise (i.e., LH). On average, the pitch rose by 22.67 Hz. On the other hand, there is no significant pitch change with the short vowel in *sekai* ‘world’ in Figure 2 (right); the average change was 8 Hz.

A similar yet distinct pattern emerges for the word-medial position. As illustrated in Figure 3 (left), a long vowel in the word-medial position is associated with a pitch drop, but not with a pitch rise.

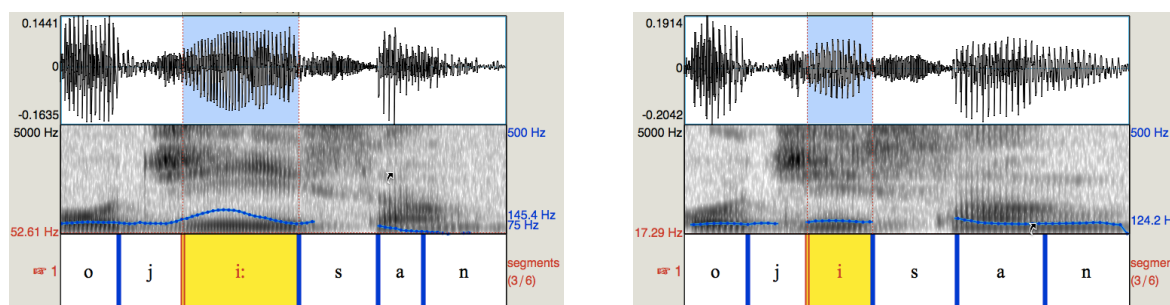


Figure 3. Length distinction in word-medial position

In the word *ojiisan* ‘grandfather’ in Figure 3 (left), the long vowel is realized as HL (i.e., pitch drop). The maximum pitch within this long vowel was averaged 168 Hz; the average minimum pitch was 112.64 Hz, resulting in a drop of 55.33 Hz. On the other hand, there is no significant pitch change with the short vowel in *ojisan* ‘uncle’ in Figure 3 (right); the average difference between the maximum and minimum pitch was 8 Hz. The crucial difference between the word-initial position and the word-medial position is that the former can be accompanied by either a pitch drop or a pitch rise, but the latter can be only associated with a pitch drop. I will discuss why such an asymmetry arises in the next section.

A completely different picture emerges for the word-final position. As illustrated in Figure 4, the pitch goes down for both the long vowel (left) and the short vowel (right), and it does so very steadily.

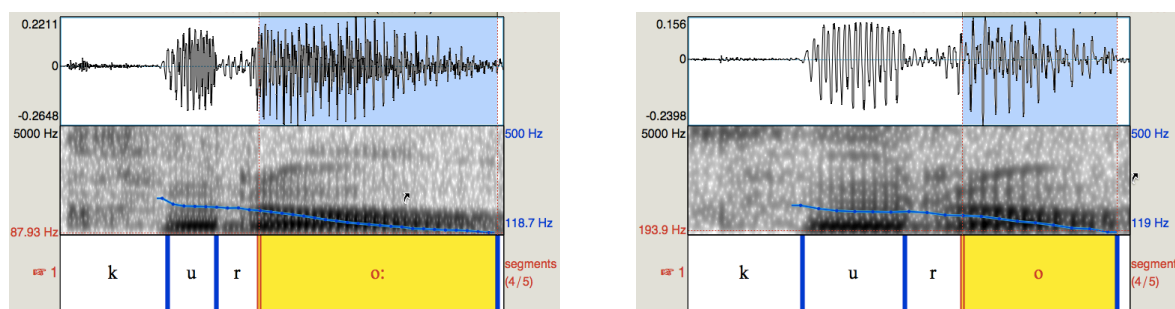


Figure 4. Length distinction in word-final position

I argue that this “decrease of pitch” must be distinguished from “pitch-drop” that we discussed above. In a pitch-drop (e.g., *ojiisan* ‘grandfather’ in Figure 3, left), two morae in a long vowel are linked with two distinct pitches: the first mora with H and the second mora with L. As a result, we observe a “bump,” where the high pitch peaks. In contrast, we observe a smooth steady decrease of pitch in the word-final position as if there were only one mora. I surmise that the two morae of a word-final long vowel are associated with a single pitch (i.e., LL or HH), and that the (gradual) decrease of pitch is due to the decrease of air in the lungs, approaching the end of an utterance. This second point is also consistent with the fact that the pitch decrease occurs with a short vowel as well in the word-final position despite the fact there is only one mora (thus only one pitch). For these reasons, I argue that long vowels in the word-final position are characterized by a “flat” pitch. What this implies is that vowels in the word-final position have a very similar, if not identical, pitch pattern whether they are short vowels or long vowels, which, in turn, implies that vowel lengths cannot be distinguished in terms of pitch in the word-final position. While this may seem to undermine my proposal, it, in fact, supports it. It has been observed that the phonemic distinctions between long and short vowels are often neutralized in word-final positions in Japanese (Kubozono, 2002). For example, in words such as *ohayoo* ‘good morning,’ *hontoo* ‘really,’ and *sensee* ‘teacher,’ the long vowels can be pronounced short, “blurring” the differences in duration. In fact, words, such as *ohayoo* and *hontoo*, are sometimes even spelled with a short vowel, which never happens with long vowels in the word-initial or word-medial positions. While native speakers of Japanese are very keen to vowel length distinctions in word-initial and word-medial positions, they are much less sensitive in the word-final position. This asymmetry can be attributed to the fact that a pitch cue is available in the former but not in the latter. The characteristic pitch contours for long vowels in different positions are summarized in Table 2.

Table 2

Summary of Characteristic Pitch Contours of Long Vowels

Position	Pitch Contour	Example
Word-initial	“Pitch drop” (HL)	<u>ko.o</u> .shi (HLL) ‘lecturer’
	“Pitch rise” (LH)	se. <u>e</u> .ka.i (LHHH) ‘answers’
Word-medial	“Pitch drop” (HL)	o. <u>ji.i</u> .sa.n (LHLLL) ‘grandfather’
Word-final	“Flat” (LL)	ku.ro. <u>o</u> (HLL) ‘hardship’
	“Flat” (HH)	ru.bi. <u>i</u> (LHH) ‘ruby’

As shown in Table 2, long vowels are associated with specific pitch contours in the word-initial and word-medial positions. I discuss in the next section that these pitch contours follow from general accentuation rules in Japanese.

RELIABILITY OF PITCH CONTOURS

I demonstrate in this section that the distribution of the characteristic contours I discussed above are predictable by general rules of accentuation although the distribution of long and short vowels itself is not predicable.

Recall that a long vowel in the word-initial position is always characterized by either a “pitch drop” or a “pitch rise.” This fact follows from an accentual rule, the “initial lowering” rule (Haraguchi, 1977). According to this rule, the first mora of a word is always low-pitched in Japanese unless it is accented. It is important to note that the accented mora is marked H, and the remaining morae are marked L. With this background in mind, first consider a situation where the first mora is accented. If the first mora is accented, then the first mora is H, and the second mora is L, creating a pitch drop. If, on the other hand, the first mora is not accented, the first mora is L, and the second mora has to be H, resulting in a pitch rise. What this means is that the pitch of the first mora and that of the second mora must be always different from each other. In other words, the first two morae of a word must be either HL (i.e., pitch drop) or LH (i.e., pitch rise).

Let us now examine the word-medial position. Recall that, unlike in the word-initial position, long vowels in the word-medial position are always characterized by a pitch drop. In other words, long vowels cannot be associated with a pitch rise in word medial positions. This observation is consistent with the accentuation rule that “special morae” can never be accented. Special morae include the moraic nasal and moraic obstruents as well as the lengthened part of a long vowel. With this background in mind, let us examine two scenarios. First, in order for a pitch drop to obtain, the first mora of a long vowel has to be H and the second mora has to be L. This pitch contour is consistent with the pitch rule since the second mora (the lengthened part of the long vowel), which is a special mora, is not accented. Second, a pitch rise occurs only when the first mora of a long vowel is L and the second mora is H. Since the second mora of a long

vowel cannot be accented because it is a special mora, the LH pitch contour (i.e., pitch rise) is never available in the word-medial position.

SUMMARY

Since vowel length is a phonemic distinction in Japanese, the distribution of long and short vowels is unpredictable. This presents serious difficulty for both learners and instructor of Japanese. In this paper, I demonstrated that, while the length distinctions themselves are not predictable, the pitch contours associated with them are predictable, by showing that they follow from general rules of accentuation.

Given the predictability of pitch contours and the unpredictability of length distinctions, I argue that it is more effective to capitalize on the pitch rather than the duration in teaching vowel length distinctions in Japanese. More specifically, I suggest that instructors draw learners' attention to the lack or the presence of a pitch contour for short or long vowels, respectively. Since learners tend to mistake long vowels for short vowels (Oguma, 2000), training learners to identify the characteristic pitch contours of long vowels is crucial. In addition, since these distinctions often disappear in word-final positions, it is best to practice vowel length distinctions in the word-initial and word-medial positions.

ACKNOWLEDGMENTS

I would like to thank the participants of the 6th Annual PSLT conference and the anonymous reviewer for their many constructive comments. Thanks are also due to the College of Humanities and Social Sciences and the Department of Modern and Classical Languages at Western Washington University for the funds made available to participate in the conference.

ABOUT THE AUTHOR

Masanori Deguchi (masanori.deguchi@wwu.edu) is an Associate Professor of Japanese and Linguistics at Western Washington University in Bellingham, Washington. He received his Ph.D. in Linguistics (major) and Philosophy (minor) from Indiana University, Bloomington in 2003. Ever since, he has been teaching at Western Washington University. He primarily teaches Japanese courses, including Japanese Phonetics, and he also teaches linguistics (e.g., phonology and semantics) and language pedagogy. His main area of expertise is in the syntax and semantics of Japanese. His representative publications include "Two Indirect Passive Constructions in Japanese" in Alexiadou and Schäfer (eds.) *Non-canonical passives*, John Benjamins.

Masanori Deguchi
Department of Modern and Classical Languages
516 High Street, Miller Hall 217E
Bellingham, WA 98226, USA
Phone: +1 360 650 4155

REFERENCES

- Fujisaki, H., Nakamura, K., & Imoto, T. (1975). Auditory perception of duration of speech and non-speech stimuli. In G. Fant & M. A. A. Tatham (eds). *Auditory analysis and perception of speech* (pp. 197–219). London: Academic Press.
- Haraguchi, S. (1977). *The tone pattern of Japanese: an autosegmental theory of tonology*. Tokyo: Kaitakusha.
- Han, M. (1962). The feature of duration in Japanese. *Onsei no Kenkyu*, 10, 65-80.
- Hirata, Y. (2004). Effects of speaking rate on the vowel length distinction in Japanese. *Journal of Phonetics*, 32, 565-589.
- Kinoshita, K., Behne, D., & Arai, T. (2002). Duration and F0 as perceptual cues to Japanese vowel quantity. In J. Hansem & B. Pellom (eds). *7th International Conference on Spoken Language Processing* (pp. 757–760), Baixas, France: ISCA.
- Kubozono, H. (2002). Temporal neutralization in Japanese. In C. Gussenhoven & N. Warner (eds). *Laboratory Phonology 7* (pp. 171–201). Berlin & New York: Mouton de Gruyter.
- Ladefoged, P., & Disner, S. F. (2012). *Vowels and consonants*. West Sussex, U.K.: Wiley-Blackwell.
- Oguma, R. (2000). Perception of Japanese long vowels and short vowels by English-speaking learners. *Japanese-Language Education around the Globe*, 10, 43-54.
- Shibatani, M. (1996). *The languages of Japan*. Cambridge: Cambridge University Press.
- Tajima, K., Kato, H., Rothwell, A., Akahane-Yamada, R., & Munhall, K. G. (2008). Training English listeners to perceive phonemic length contrasts in Japanese. *Journal of Acoustic Society of America*, 123(1), 397-413.
- Toda, T. (2002). *Second language speech perception and production: acquisition of phonological contrasts in Japanese*. Maryland: University Press of America.
- Ueyama, M. (2000). *Prosodic transfer: An acoustic phonetic analysis of Japanese-accented English*. Ph.D. dissertation, University of California at Los Angeles.
- Vance, T. (2008). *The sounds of Japanese*. Cambridge: Cambridge University Press.

APPENDIX: Pitch Measurements

Kooshi 'lecturer'

	Recording 1	Recording 2	Recording 3	Average
Max	155	159	161	158.33
Min	115	113	114	114.00
Difference	40	46	47	44.33

Koshi 'old paper'

	Recording 1	Recording 2	Recording 3	Average
Max	164	153	153	156.67
Min	158	139	139	145.33
Difference	6	14	14	11.33

Seekai 'correct answer'

	Recording 1	Recording 2	Recording 3	Average
Max	146	149	143	146.00
Min	120	124	126	123.33
Difference	26	25	17	22.67

Sekai 'world'

	Recording 1	Recording 2	Recording 3	Average
Max	160	163	154	159.00
Min	150	155	148	151.00
Difference	10	8	8	8.00

Ojiisan 'grandfather'

	Recording 1	Recording 2	Recording 3	Average
Max	169	167	168	168.00
Min	117	107	114	112.67
Difference	52	60	54	55.33

Ojisan 'uncle'

	Recording 1	Recording 2	Recording 3	Average
Max	128	126	126	126.67
Min	118	122	121	120.33
Difference	10	4	5	6.33

A NAIL IN THE COFFIN OF STRESS-TIMED RHYTHM

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

Stress-timed rhythm as applied to English (Pike, 1945) stands on three pillars. Research has convincingly invalidated two of these, undermining the claim that English is an exemplar of stress timing. If not stress timed, then what rhythm does English exhibit? This report describes a rhythm pattern that is not only widely attested in actual usage but is also simpler than the discredited pattern. Furthermore, learners are able to use it for clearer spontaneous speech.

INTRODUCTION

Prosody—rhythm and intonation—is a prime of spoken English; it gives the speech stream its sound shape (Bolinger, 1981). When the prosody of a phrase is right, a myriad of details of the language align to convey one’s meaning without distraction; segmental errors can even pass unnoticed (Kjellin, 1999). When the prosody is wrong, the consequences for communication can be serious. This is why prosody is central to intelligibility (Derwing & Rossiter, 2003; Hahn, 2004). Recognizing this, many TESOL practitioners focus on developing learners’ prosodic skills.

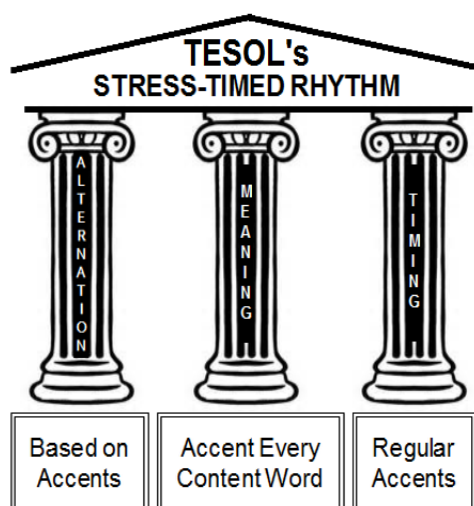
The questions we raise here are: In the area of rhythm, are we teaching the right content? If not, then what should we teach? Not long ago, these were pressing questions for this author, a TESOL practitioner and researcher. My reflections on a journey of loss and discovery are included.

THE CANON

Seventy years ago, when our field was young, insights moved from linguistics to ESL textbooks with little delay because most linguists were also language teachers. So when the linguist, Kenneth Pike, distinguished different types of rhythm, calling one “stress timed” and the other “syllable timed” in 1945, these distinctions spread quickly to the new profession (Prator, 1951; Lado & Fries, 1954).

Before long these ideas took root solidly in our TESOL narrative about the rhythm of English (stress-timed) and of most other languages (syllable-timed). They are now found in virtually every ESL pronunciation textbook, including mine (Dickerson, 1989/2004; Hahn & Dickerson, 1999a, 1999b). These rhythm types have the status of established fact, even though at the start they were untested hypotheses.

All of the features now associated with stress timing in TESOL were present in the 1950s: levels of stress—accented vs. unaccented,¹ the relative meaningfulness of English words—content words vs. function words,² and timing—a regular occurrence of accents. Although opinion was split at first about whether every content word should be stressed, the consensus came to support this view.³ The results are the three pillars of stress-timed rhythm in the TESOL community.



THE TRAUMA OF LOSS

Although Pike uses stress timing and syllable timing to describe different rhythmic features of English, he also comments that syllable timing is dominant in Spanish (1945:35). Pushing this idea further, Abercrombie (1967) reasoned that, since the production of the two rhythm types requires mutually exclusive articulatory processes, languages should therefore exhibit one type of rhythm or another. His hypothesis explicitly addresses the third pillar above: accents come at regular intervals; it says nothing about the location of accents—the second pillar. His hypothesis challenged investigators to categorize languages according to their rhythmic pattern. It was

¹ An accent is a heavy stress accompanied by a pitch change.

² Pike coined the terms “content word” and “function word.” For him, content words are broader than “lexical words”—nouns, main verbs, adjectives, adverbs (of time, place, and manner). They also include demonstrative and indefinite pronouns, interrogatives, and interjections (Pike, 1945:118). All others are function words.

³The two dominant pronunciation textbooks at the time took different positions. Prator’s 1951 text promotes stressing every content word; Lado & Fries (1954), following Pike’s work, promotes deaccenting of content words. There are many reasons that Prator’s position gained the ascendancy besides the fact that his text stayed in print 20 years longer than Lado & Fries’ text.

disappointing work; for two decades researchers were unable to find a single instance of a stress-timed or syllable-timed language. Cauldwell (2002:1) summarizes:

The evidence from research is overwhelmingly against the hypothesis that languages are either ‘stress-timed’ or ‘syllable-timed’. It is not possible to divide language into either ‘syllable-timed’ or ‘stress-timed’ categories; it is not the case that stresses occur at equal time-intervals in ‘stress-timed’ languages; it is not the case that syllables occur at equal time intervals in ‘syllable-timed’ languages; so-called ‘syllable-timed’ and ‘stress-timed’ languages are alike in having variations in syllable-length; so called ‘syllable-timed’ and ‘stress-timed’ languages are alike in having variations in inter-stress-interval length.⁴

My reaction to this conclusion was like that of many of my colleagues: Although the evidence was indeed overwhelming, it was still hard to drop stress and syllable timing. In the first place, the dichotomy offered a way to talk about differences I felt existed.⁵ At first I wanted to keep talking about stress timing but with enough qualifiers to cover my bases—referring to “tendencies,” “relatively regular pace,” “stress-based rhythm” (Dauer, 1983)—all of which actually reject the claim of regularity in timing. In two contexts, however, this “fudge language” did not feel right: I did not want to continue telling an untruth to my MATESL students now that I knew better. Also in my ESL classes, I could no longer justify insisting that learners keep pace with a rhythm that was not really there. However, the biggest deterrent to letting go of stress timing was that I had no idea how to describe English rhythm if I did let go.

It took some time to realize that holding on to the myth of stress timing in any form would undercut any incentive to explore the reality of English rhythm. Ultimately, a determination emerged to find out what is going on with rhythm.

⁴The evidence against stress timing: Instead of phrases being compressed or expanded to some uniform size, the length of inter-stress intervals is proportional to the number of syllables between the stresses (Roach, 1982; Dauer, 1983; Halliday, 1994; Bertrán, 1999). As early as 1939, Classe had tested the hypothesis, as articulated by Jones (1918), and found that the conditions necessary for strict stress timing to appear in everyday English speech are so rare as to occur only by accident (Cauldwell, 1996).

⁵We now know that perception and production are asymmetrical; up to a certain threshold of variability, we perceive more regularity in speech than is actually present (Lehiste, 1977, 1979). In fact, an irregular rhythm is functional; it keeps the listener attending to the message and not the tempo (Cauldwell, 2002).

THE DELIGHT OF DISCOVERY

When I went in search of the real nature of English rhythm, I was surprised to discover that a number of scholars already had an answer. Even more surprising, they agreed on the kind of rhythm English has. Furthermore, they had been telling us about it for years. We TESOL professionals, however, were so confident about how English rhythm works that we had stopped paying attention. Three phonologists represent this group of researchers. They describe their conclusion about English rhythm using their own, but equivalent, terminology.

By 1961, the American linguist, Dwight Bolinger, had figured out the basic shapes of English rhythm. His metaphor of a suspension bridge represents a phrase, or what he calls a “melodic group,” with two accents—the two towers. He also describes a phrase with a single accent, but does not provide a graphic. If he had, it would look like the right-hand image below.



About this model of rhythm, Bolinger (1961) says,

...from the standpoint of running frequency they [melodic groups containing two accents and melodic groups containing one accent] probably outnumber all others in everyday conversational statements. (p. 135)

Bolinger is describing “everyday conversational statements”—the one- and two-second bursts of speech with which we all talk, commonly known as spontaneous speech.

Richard Cauldwell, a British scholar, also studies spontaneous speech. He cites frequency counts of “tone-units” in three studies and concludes (Cauldwell, 2002):

The majority of speech can be divided up into tone-units of three sizes: single-prominence, double-prominence, and triple prominence tone-units... For all three [spoken] texts, single and double prominence tone-units account for very close to 90% of all [1700] tone units. (pp. 8, 15)

Phrases that have three or more accents are rare. They also tend not to be spontaneous speech. They represent instead “the successful delivery of a preplanned ‘chunk’ which may have been uttered before” (p. 17), snippets of verse, or “idiomatic, or semi-idiomatic material” (p. 18).

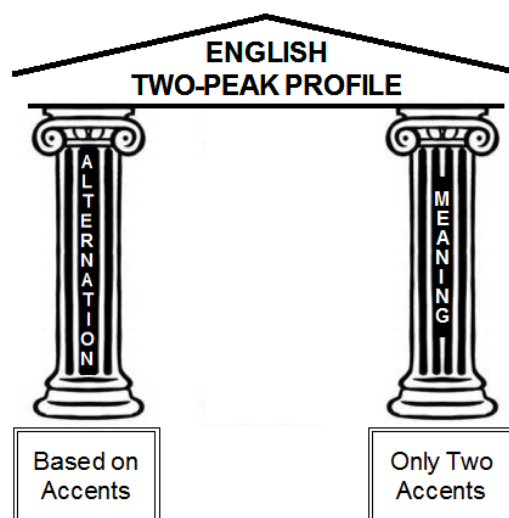
Another British researcher, John Wells, makes the same claim about naturally occurring ‘intonational phrases’ (IP) (Wells, 2006):

An IP usually contains only one or two accents (onset and nucleus, or just a nucleus). (p. 192)

He notes that phrases with more than two accents are uncommon and are found more “in scripted material and in material read aloud” than in spontaneous conversation. Spontaneous phrases are short and contain few accents because that small size is “the basic chunk for mental planning” (p. 192).

In short, whether referring to American English or British English, these scholars are telling us that a well-attested alternative to stress timing does exist. It is real, not hypothetical.⁶ Like stress timing, the alternative is based on an alternation of what we call peaks (each with a single accent) and valleys (each with one or more lightly stressed or unstressed syllables); the first pillar (above) stands. Unlike stress timing, a spontaneous phrase has at most only two peaks according to the alternative model. If the phrase has more than two content words, then some of the content words will not carry a peak. Furthermore, none of these scholars considers timing to be at all relevant. These facts topple the second and third pillars of stress timing as applied to English.

It may seem strange to speak of rhythm without timing. However, this is what the research shows. Uniformity in the occurrence of peaks is not inherent in English (see footnote 4). Instead, rhythm is to be found in the repeated appearance of a pattern in the alternation, namely, two meaningful accents that surface in phrase after phrase. We call this the **two-peak profile**.



ACCESSING THE MODEL

If we accept that the two-peak profile is at play in everyday spoken English, then the practical question is: How do we introduce this new model to ESL learners?

⁶For spectrographic profiles of phrases with two accents see Dickerson, 2011, pp. 75-76.

Of the two peaks, we already know a lot about the accent called variously the nucleus, the tonic, the focal stress.⁷ It is the single accent in one-peak phrases, and the second accent in two-peak phrases. In our ESL materials, it is the **primary stress** (●). By contrast, linguistic research tells us little about the first accent in two-peak phrases, generally known as the onset. In our teaching materials, we call it the **anchor** (○). If the learner cannot identify the anchor, it is impossible to create an authentic two-peak profile. So where is it?

At the University of Illinois we have been working on this question for several years. We have benefitted from having available transcripts from other linguists in which the anchor and primary stress have been marked on thousands of phrases in a discourse context. Together with our own data, we have pushed the analysis far enough to allow us to develop teaching materials for university-level learners. It has been encouraging to see that these learners understand. They can quickly make accurate predictions, and the rhythm they produce sounds like English rhythm. The second edition of our textbook, *Speechcraft*, is built around our research and what we have learned from trialing these materials with learners (Dickerson & Hahn, in press).

NAILING DOWN THE ANCHOR

Our anchor-placement strategy rests on five guiding positions—two goals and three procedures. First, our objective is to define the anchor in spontaneous speech, not in written texts delivered orally. The structure and length of phrases in the two styles of speech are radically different. The two-peak profile is native only to unrehearsed speech, the kind filled with short phrases limited by the speaker's and listener's natural on-the-fly processing powers.⁸ Second, we aim to predict the anchor where it does not draw special attention to itself, in the neutral position. We discuss using the anchor for emphasis later. Third, we always start with the primary stress in place: Find the primary stress; then find the anchor. Fourth, we scan for the anchor at the beginning of the phrase and move to the right. Finally, we look for particular parts-of-speech.

Wells provides a preliminary hypothesis: "In unemphatic speech, only the *first* content word [left of the nucleus] receives an accent" (2006:207). Our test of this hypothesis includes these phrases.

⁷Addressing ESL/EFL teachers, John Wells offers a generous, yet easily accessible, account of the nucleus (2006:93-186). A pedagogical counterpart for ESL/EFL learners can be found in Hahn & Dickerson (1999a:62-79; 1999b:80-81).

⁸In our pedagogical materials, we call these phrases **message units**. They are equivalent to what Bolinger calls a "melodic group," what Cauldwell calls a "tone unit," and what Wells calls an "intonational phrase." In structure, they typically correspond to a grammatical unit and therefore express a coherent thought. In length, they are generally between one and seven words in length. Orally, they are most often identifiable by a pause and a phrase-final intonation pattern within which we find the primary stress. All of the following examples are instances of message units.

○ ●

September's quite a full month.

○ ●

We're attending a conference in Santa Barbara.

○ ●

It's a really nice break for both of us.

While only the first sentence conforms to the hypothesis, we still learn something important: The anchor likes **nouns**. It will hop over a verb and a variety of modifiers to reach a noun.

The phrases above also seem to show that the anchor does not like modifiers much. That generalization, however, is too broad, as the next phrases prove.

○ ●

And we're really looking forward to it.

○ ●

Maybe we'll even rent a car.

The anchor does not skip over all modifiers. If they modify a verb or a whole phrase, then the anchor seems to like **adverbs** fine.

Verbs, though, appear to be quite out-of-favor. The anchor hops over the verb in *We're attending a conference in Santa Barbara*. The following phrases refine the observation.

○ ●

Take a look at the map again.

○ ●

Let's visit Sequoia National Park.

For a verb to attract the anchor, it has to be either a second person imperative or a first person imperative. So **imperative** verbs attract the anchor.

To this point, the anchor seems to attach to some content words but not others. How does it like loud function words, namely, interrogative and negative words and demonstrative and indefinite subject pronouns?⁹ The following sentences provide a clear answer.

⁹In our pedagogical materials, words divide into three categories—content words, *loud* function words, and *soft* function words. Loud function words have the same potential for peak

○ ●

Where would you like to camp?

○ ●

He never packs a radio.

○ ●

That's the trail I was talking about.

○ ●

Did anyone remember the insect repellent?

Loud function words are solidly in the anchor camp.

On its first pass through a phrase, the anchor stops on a Noun, Adverb modifying a verb or whole phrase, Imperative verb, and any type of Loud function word, whichever comes first in a phrase. The acronym for this first pass is **NAIL**.

Of course, the first pass could yield no NAIL word left of the primary stress. In this case there are two possibilities. First, although there is no NAIL word, there may still be another content word as in these comments about giant sequoias.

Hiking through a small grove of them, |

○ ●

we were truly speechless.

○ ●

They're as majestic as they are ancient.

The anchor lands on the first content word in the phrase: **1st CW**.

The second possibility is that there is no content word left of the primary stress. If not, the phrase has no anchor at all: \emptyset .

•
I can believe it!

stress that content words have—nouns, adjectives, verbs, and adverbs. Soft function words usually carry valley stress—either tertiary or unstressed.

•

You should have warned me.

With the basic framework in place, we turn to the last category of anchor-attracting words we call “number.” It encompasses several related groups. The examples above demonstrate that the anchor hops over most modifiers on the first pass, particularly adjectives, but not all adjectives.

○

•

With only three days to sightsee, | [how far can we go?]

○

•

Fifty percent of the canyon | [is closed to tourists.]

The anchor clearly prefers a cardinal number if it encounters it first. Not surprisingly, the anchor reacts the same to ordinal numbers.

○

•

His first caution was about the bears.

Ordinal numbers are members of a larger group of “order” words, like *next*, *last*, *previous*, *following* as in sentences such as these:

○

•

Our last hike was to Half Dome.

○

•

The following dates are still open | [for a ranger tour.]

Order words also include superlatives—the *-est* form and adjectives and adverbs preceded by *most*. Superlatives identify the extreme end of some order or continuum.

○

•

The largest sequoia is the General Sherman.

○

•

The most challenging ascent for climbers | [is the ‘Nose’ of El Capitán.]

The anchor is attracted to either the single-word superlative or the adjective or adverb of the paraphrastic superlative, not the word *most*.

If we apply NAIL to these phrases with number and order words, we will mispredict the anchor on the first noun. To avoid this error, the anchor elects the first number or order word it encounters in the first pass. We include this category as “#”: #NAIL.

EMPHATIC USES OF THE ANCHOR

Neutral: The evening was absolutely wonderful. That's not what I meant.

Emphatic1: The evening was absolutely wonderful. That's not what I meant.

Emphatic2: ○ ○ ● ○ ○ ●

The evening was absolutely wonderful. That's not what I meant.

CONCLUSION

1. Learners' rhythm will sound like *English* rhythm when learners use the *actual* rhythm of English as a model instead of an unnatural, artificial model. Specifically:

192

- English speakers do not accent every content word and loud function word in unemphatic phrases.
 - The consensus of research is that accents do not come at equal intervals in spoken English.
 - An authentic alternative to stress timing is available in the two-peak profile.
2. Listeners expect to hear the two-peak profile and can decipher messages with an expected rhythm more easily than with an unnatural rhythm (Kjellin, 1999:11-12). Specifically:
- Stress timing generates peaks where valleys should be, thereby reducing the perceived prominence of the primary stress (Dickerson, 2011:75-78).
 - Stress timing, when ‘correctly’ implemented, sends a strongly emphatic meaning in phrases containing multiple content and loud function words, a meaning which the learner is usually unaware of and does not intend.

As difficult as it is to say goodbye to stress timing, there is something quite powerful and authentic waiting for learners when we do let it go. We would like to think that this research may help put a NAIL in the coffin of stress timing.

ACKNOWLEDGEMENTS

The author wishes to thank Laura Hahn, Ester Cha, Yelena Tower, and Lonna Dickerson for their helpful feedback on this paper.

ABOUT THE AUTHOR

Wayne Dickerson is professor emeritus in the Department of Linguistics at the University of Illinois at Urbana-Champaign where he taught courses in English phonology (online and face-to-face) and ESL pronunciation. His research focuses on pedagogical applications of phonetics and phonology, pronunciation pedagogy, the value of orthography for learners, phonological variability, and pronunciation assessment. His two pronunciation textbooks are *Stress in the Speech Stream: The Rhythm of Spoken English* (1989), 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.

Contact information:

Wayne Dickerson
7 Hale Haven Court
Savoy, IL 61874
dickerson@illinois.edu

REFERENCES

- Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh: Edinburgh University Press.
- Bertrán, A. (1999). Prosodic typology: On the dichotomy between stress-timed and syllable-timed languages. *Language design*, 2, 103-130.
- Bolinger, D. (1961). Three analogies. *Hispanica*, 44, 135-136.
- Bolinger, D. (1981). *Two kinds of vowels, two kinds of rhythm*. Bloomington, IN: Indiana University Linguistics Club.
- Cauldwell, R. T. (1996). Stress-timing: Observations, beliefs, and evidence. *Eger journal of English studies*, 1, 33-48.
- Cauldwell, R. T. (2002). The functional irrhythmicality of spontaneous speech: A discourse view of speech rhythms. *Apples*, 2, 1-24.
- Classe, A. (1939). *The rhythm of English prose*. Oxford: Blackwell.
- Dauer, R. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of phonetics*, 11, 51-62.
- Derwing, T. M., & Rossiter, J. M. (2003). The effects of pronunciation instruction on the accuracy, fluency, and complexity of L2 accented speech. *Applied language learning*, 31, 1-18.
- Dickerson, W. B. (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. B. (2011). Upstream destressing: Another step toward natural speech. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 2nd pronunciation in second language learning and teaching conference*, Ames, IA (pp. 70-81). Ames, IA: Iowa State University.
- Dickerson, W. B., & Hahn, L. D. (in press). *Speechcraft: Discourse pronunciation for advanced learners*. (2nd ed.). Ann Arbor: The University of Michigan Press.
- Hahn, L. D. (2004). Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL quarterly*, 38(2), 201-223.
- Hahn, L. D., & Dickerson, W. B. (1999a). *Speechcraft: Discourse pronunciation for advanced learners*. Ann Arbor: The University of Michigan Press.
- Hahn, L. D., & Dickerson, W. B. (1999b). *Speechcraft: Workbook for academic discourse*. Ann Arbor: The University of Michigan Press.
- Halliday, M. (1994). *An introduction to functional grammar*. (2nd ed.). London: Edward Arnold.
- 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.D. Joseph, & B. Palek (Eds.), *Proceedings of LP '98 (Linguistics and phonetics conference)*, Ohio State University, Columbus, OH (pp. 373-398). Prague: The Karolinum Press.
- Lado, R., & Fries, C. (1954). *English pronunciation*. The University of Michigan Press.
- Lehiste, I. (1977). Isochrony reconsidered. *Journal of phonetics*, 5, 253-263.

- Lehiste, I. (1979). The perception of duration within sequences of four intervals. *Journal of phonetics*, 7, 313-316.
- Pike, K. L. (1945). *The intonation of American English*. Ann Arbor: The University of Michigan Press.
- Pike, K. L. (1942). Pronunciation. In *Intensive course in English for Latin-American students*. Vol 1. Ann Arbor: The University of Michigan Press.
- Prator, C. L., Jr. (1951). *Manual of American English pronunciation for adult foreign students*. Los Angeles: University of California Press.
- Roach, P. (1982). On the distinction between 'stress-timed' and 'syllable-timed' languages. In D. Crystal (Ed.), *Linguistic controversies*, 73-79. London: Edward Arnold.
- Wells, J. C. (2006). *English intonation: An introduction*. Cambridge: Cambridge University Press.

Henrichsen, L., & Stephens, C. (2015). Advanced adult ESL students' perspectives on the benefits of pronunciation instruction. In J. Levis, R. Mohammed, M. Qian & Z. Zhou (Eds). *Proceedings of the 6th Pronunciation in Second Language Learning and Teaching Conference* (ISSN 2380-9566), Santa Barbara, CA (pp. 197-205). Ames, IA: Iowa State University.

ADVANCED ADULT ESL STUDENTS' PERSPECTIVES ON THE BENEFITS OF PRONUNCIATION INSTRUCTION

[Lynn Henrichsen](#), Brigham Young University

[Chirstin Stephens](#), Brigham Young University

What do ESL students perceive as the benefits of pronunciation instruction? The answer to this important question is unclear because very few studies on L2 pronunciation teaching have focused on students' perspectives regarding its value. This paper reports on research that replicated and extended Levis, Link, and Sonsaat's (2013) study that investigated students' perspectives on the effects of pronunciation instruction. Participants were 12 university students in an advanced English pronunciation course. The quantitative analysis showed that the students' comprehensibility and accentedness did not change significantly. Nevertheless, despite this apparent lack of progress, the qualitative analysis revealed that students found the course beneficial because it increased their awareness (of general articulatory processes and their own difficulties), built their confidence, improved their listening, and provided them with valuable pronunciation-improvement strategies.

INTRODUCTION

In the history of research on L2 pronunciation teaching, many studies have focused on a variety of issues. For instance, researchers have investigated factors that affect L2 pronunciation accuracy (Piske, MacKay, & Flege 2001; Purcell & Suter 1980; Suter 1976), student achievement of pronunciation course objectives under different instructional conditions (de Bot & Mailfert, 1982; Derwing, Munro, & Wiebe, 1998; Macdonald, Yule, & Powers, 1994; Yule, Hoffman, & Damico, 1987; Yule & Macdonald, 1995), native and non-native English speakers reactions to English learners' pronunciation (Hahn, 2004; Riney, Takagi, & Inutsuka; 2005), and L2 learners' perceptions of their pronunciation needs and strategies (Derwing & Rossiter, 2002).

Nevertheless, very few studies have focused on an important perspective on pronunciation instruction—that of the students receiving it. Exceptions to this trend are studies by Kang (2010), who looked at ESL learners' expectations of pronunciation lessons and their attitudes toward the variety of English spoken by their instructors, and Levis, Link, and Sonsaat (2013), who investigated advanced-proficiency, older learners' confidence resulting from pronunciation instruction. This paper reports on research designed to help fill that gap by replicating and extending Levis, Link, and Sonsaat's (2013) study.

Research Questions

Our research questions were the following:

1. Did the pronunciation course result in measurable improvement in the students' comprehensibility and decrease the accentedness of their speech?
2. How did students' attitudes, knowledge, and confidence regarding English pronunciation change over the course of the semester?
3. What pronunciation-improvement and communication strategies did students learn and use outside of class during this course?

METHODS

Participants

The participants were 12 adult students (5 males, 7 females) enrolled in a university-based advanced English pronunciation course that met three times a week for a 15-week semester. They came from various L1 backgrounds—Spanish (6), Korean (3), Chinese (1), German (1), and Japanese (1). While most were university students, some were older professionals desiring to improve their pronunciation. All had chosen to enroll in this elective ESL pronunciation course, which focused on both segmentals and suprasegmentals in communicative contexts and was taught by a TESOL professor and two undergraduate TESOL interns.

Procedures

The course began with a diagnosis of each individual student's pronunciation, which resulted in a "Personal Pronunciation Improvement Plan" (or "prescription" [Rx]). Course activities included speeches, articulatory explanations and practice, *Pronunciation Matters* (Henrichsen, Green, Nishitani, & Bagley, 1999) units and associated activities (e.g. storytelling, peer practice cards, etc.), poetry, Jazz Chants, video voiceovers (Henrichsen, 2015), and reader's theatre.

Instruments and Analysis

The measures used in this research were both quantitative and qualitative. The quantitative measures assessed participants' degrees of accentedness and comprehensibility as manifest in course-initial and course-final audio recordings. For these recordings, the participants read aloud Prator and Robinett's (1985) accent inventory. Both the accentedness and comprehensibility of each ESL student's pre- and post-course speech samples (presented in random order) were later rated by seven TESOL MA students using a nine-point Likert scale (Derwing & Munro, 1997; Munro & Derwing, 1995).

The qualitative analysis involved students' written responses to ten open-ended questions (given as an end-of-semester "final exam," see Appendix A) based on Levis, Link, and Sonsaat (2013). Participant responses were analyzed by two independent raters using grounded theory (Strauss & Corbin, 1990). Key points that emerged from the data were grouped into themes noted by both raters.³

³ Because all data collection was done as part of instructional activities that were a regular part of the class, the study was determined to be exempt from informed consent regulations. Nevertheless, the research protocol was still submitted to the university's Institutional Review Board and subsequently approved.

RESULTS

Quantitative

The quantitative analysis showed that students' comprehensibility and accentedness did not change significantly over the course of the semester.

The overall class average comprehensibility rating was 4.04 at the beginning of the course, and the post-course rating was 4.03. This very slight change—even though it moved in the direction of improved comprehensibility—indicated that at the end of the course the learners were not significantly easier to understand. The results of a *t*-test run on these data showed a probability of .99 that the two means would not be significantly different if the test were repeated. A calculation of effect size (with Cohen's *d* equaling .003) further confirmed the conclusion that the two means were not different.

Likewise, the average accentedness rating for all students was 5.12 at the beginning of the course and 5.10 at the end. This slight drop indicated a general, but very slight, move toward being less accented. In this case, the *t*-test results showed a probability of .96 that the means would not be different if the tests were repeated. Confirming this conclusion, the effect size calculation resulted in a Cohen's *d* of .01.

These findings, which were not shared with the students, were consistent with the results of numerous previous studies on instruction and pronunciation improvement (Levis, Link & Sonsaat, 2013; Macdonald, Yule, & Powers, 1994; Madden, 1983; Purcell & Suter, 1980; Suter, 1976). In fact, results like these led many people to give up on pronunciation instruction a few decades ago (Morley, 1991). Nevertheless, the message of our research findings is a different one—Don't give up! Changing learners' ingrained, subconscious pronunciation behaviors is a long, complex process. In a relatively short period of time (one 15-week semester), learners often do not make enough progress for the improvement in their pronunciation to be statistically significant—especially if they are advanced-level learners who have been using English for years and have passed their “window of maximal opportunity” (Derwing & Munro, 2014). Also, lack of measurable improvement may be due to problems with instrument validity or sensitivity.

Qualitative

Despite students' apparent lack of progress, the qualitative analysis revealed that they still found the course beneficial. Overall, four general themes emerged. The course increased students' (1) awareness, (2) confidence, and (3) listening comprehension, and students also valued (4) the pronunciation-improvement (and communication) strategies that they learned. Nevertheless, this overall analysis masked a number of more specific yet important points since many of the questions examined different aspects of the course and led in different directions. Therefore, we conducted ten qualitative analyses, one for each question. The themes that emerged from the analysis of student responses to each question are discussed below.⁴ Themes are listed in frequency order within each question, from the most frequently mentioned to the least. Actual frequencies are indicated parenthetically “(N=f).” Selected, representative student comments follow each theme.

⁴ Because of space limitations, responses to questions 7 and 9 (which are not so relevant to the focus of this paper) are not discussed here. Please contact the authors if you are interested in them.

1. In what ways was this pronunciation class *valuable* to you? In what ways do you feel you have *improved* this semester because of this class?

The two most frequently mentioned themes that emerged from the students' responses had to do with increased awareness. The first was increased articulatory awareness (N=7), as shown in this comment: "I was able to improve my pronunciation by learning how to make certain pronunciations...I trained my mouth muscles [sic]⁵ to make those pronunciations more naturally." One student noted that it was helpful for him to learn "...how to pronounce specific vowel with specific mouth shape..."

Students also increased their awareness of their individual pronunciation problems (N=6). One example of this response type was, "I have been studied mostly grammar, writing, reading in Korea. But through this class I could know the actual problem I have about English pronunciation."

Some students felt that the class had been valuable because they perceived pronunciation improvement (N=4). One student said, "My pronunciation has improved with this class because I learned many things that helped to communicate better. Such as the 's' and 'z' sounds."

2. Did this class *meet your expectations* of what a pronunciation class should be? Why or why not? What else would you have included in the class?

The dominant theme that emerged from the analysis of responses to this question was that the class had met students' expectations. An overwhelming majority (10 of 12) felt that it had. One wrote, "Yes, this class was amazing for me because I learned a lot. I was so excited to be in the class." Another response was, "I learned more than I thought I could."

This overwhelmingly positive response was very encouraging. The responses to the follow-up "Why or why not?" question were enlightening. Three students said the class met their expectations because it helped them improve, and three students connected their satisfaction with the class to increased confidence. One student said, "This class...makes me feel more confident when speaking English." Three students mentioned the help that was provided by the teaching assistants. One wrote, "I felt that she understand me and really wanted me to learn and improve. That make me feel more confident to speak more and ask more questions."

3. In what ways do you *practice* your English pronunciation (in speaking or listening mode) *outside* of ESL 302 class sessions?

As might be expected, the thing that students reported doing out of class most frequently was speaking with other people (N=7). One student said, "I had a lot of native English speaking friends outside this class. They helped me regularly to improve my speaking and especially my vocabulary."

Students also said they practiced their pronunciation outside of class reading various materials aloud (N=6), using strategies learned in class (e.g. "tracking," speaking with a mirror, or speaking in slow motion) (N=5), asking native speakers to correct their pronunciation (N=3), using music (N=3), studying with a tutor or "study buddy" (N=2), and paying close attention to problems and correcting them (N=1).

⁵ All student responses are reported here exactly as they were written, without correcting spelling, grammar, or word choice.

4. Do you still use any of the *pronunciation-improvement strategies* we talked about earlier in the semester? Which ones? How often? When?

Class instruction early in the semester focused on a variety of pronunciation improvement strategies, and our third research question focused on students' continuing use of them. In their responses to the preceding "final exam" question, five students mentioned using these strategies. Question 4 asked them which strategies students used and when.

Reading aloud emerged as the most frequently mentioned strategy. Almost half of the responses (N=5) mentioned it. One said "At home I choose from my prescription and I read to my kids, they loved when I read to them they have fun helping me out."

Three students mentioned tracking, saying it was "really helpful" and "useful" while two others mentioned using music. One said, "In my kindle I have Pandora's radio and I can sing a song because I can read the words." Other strategies mentioned by only one student each included repeating difficult words aloud, focusing on specific problems, setting personal goals, using kinesthetic devices, asking friends for correction, practicing with the textbook resources, and scheduling frequent practice sessions.

Although these responses were varied, we considered them encouraging. They showed that students had learned to take responsibility for their own pronunciation improvement as recommended by pronunciation experts (Morley, 1991). That meant that they were likely to continue improving even after the class ended.

5. How often do you speak English with others (native English speakers and non-natives) each day? In what situations? Do you speak English more often now than you did before taking this class?

Ten of the twelve students responded in a way that fit the theme "Very often." These students reported speaking with native English speakers frequently in various settings.

Encouragingly, four students reported that they spoke with native English speakers "more now than before." One student said, "I speak English more often now than I did before taking the class because I am more confident speaking English now."

6. How *confident* are you in speaking English with native speakers? Why?

The most frequent response to this question was that students were "pretty confident" (N=6). One felt pretty confident when speaking about familiar topics, and another was not afraid of starting a conversation with native speakers anymore. That student noted, "Through this class, I was able to identify my problems with English pronunciation and to train my mouth muscles to produce those sounds that I had problems with. This experience made me not afraid of talking to native speakers."

Three students reported feeling only a bit more confident, but they held out hope for the future. One said, "I feel a little bit more confident. But I need to work more. I need to learn how native speakers express themselves and find more people to practice with. I think the best way to gain more confidence is practicing and practicing and practicing."

8. In what ways do you feel your *listening* has improved during the class? Do you hear English speakers more effectively now?

Student responses revealed a perceived relationship between listening and pronunciation instruction. Apparently, for the majority of students in this class, the instruction they received in the pronunciation of English segmentals and suprasegmentals helped them perceive these sounds more accurately (N=7), validating Gilbert (1987). One said, “Yes. Understanding how pronunciation works helps me to understand when I hear English.” Another noted increased understanding due to learning to pay attention to intonation and stress.

Three students felt their listening had improved “a lot,” bringing the total number of positive responses to 10 out of 12. One student said, “I think I have improved a lot when listening comes. I feel more confident understanding what people say. Learning the different sounds and how to identify them makes me feel confident.” Another student felt that from the beginning of the class to the end of the class, her listening comprehension went from 50% to 90%. Yet another student noted, “After this class I can hear well. I can understand pronunciation with a different ear.”

10. What things that you learned in this class this semester will you *continue to use in the future*?

Although student responses to this question varied widely, a leading theme that emerged from our analysis was the importance of suprasegmentals (N=4). One student wrote, “I learned super-segmental a lot and I recognize the improvement of intonation or other things. They are relatively easier to fix. So, when I speak I can be more careful about super-segmental.” Another noted, “I will keep working on intonation, stress... For me I think is the most important problem to overcome.”

Several students (N=4) mentioned that they would persistently keep on practicing English and asking for correction to their pronunciation.

Along with developing greater confidence, some students learned not to worry so much about their pronunciation (N=3). One said, “I will use the technique of not worrying so much about myself.... I’ve learned that what is important is to continue and have confidence on myself.”

Some students were committed to continuing self-study using the instructional materials they learned about in this course (N=3).

One very positive student summed up the value of the course by saying that she would continue to use “everything.”

DISCUSSION

These findings have a number of implications for L2 pronunciation teaching. Foremost, even though measurable improvement may not be immediately apparent (due to study length, instrument choice, or other factors), students may still see a pronunciation course as valuable for various reasons, including increased awareness, improved listening skills, motivation, and perceived improvement. Additionally, the findings of this research indicate that the benefits of pronunciation instruction may extend beyond the end of a course—especially in terms of confidence, self-awareness, and strategies for continuing improvement.

This research is not without limitations. Chief among them are the small sample size and the potential for bias in the qualitative analysis. The quantitative analysis may have also suffered from instrumentation difficulties; a narrative task rather than reading a diagnostic passage aloud might have produced different results. These limitations can be overcome only by additional, similar studies. Another limitation to the generalizability of this study is that it was conducted in an ESL setting; English language learners in EFL settings would likely give different responses to many of the questions. In sum, only by comparing the results of future studies involving more students at different levels and in different settings will we be able to draw solid conclusions regarding the student-perceived benefits of pronunciation instruction.

Despite these limitations, the study reported here offers encouragement and reassurance to students in, and teachers of, ESL pronunciation classes—especially when improvement is not immediate or significant. Real pronunciation improvement is a long, complex, and difficult process. Nevertheless, our results indicate that formal instruction in pronunciation—particularly when it is communicative, enjoyable, and focused on individuals’ actual difficulties; includes strategy instruction; and places responsibility for improvement on students’ shoulders—is a valuable and beneficial from the students’ perspective.

ABOUT THE AUTHORS

Lynn E. Henrichsen is a professor of TESOL in the Linguistics and English Language Department at Brigham Young University. He teaches courses in instructional methods (including the teaching of listening, speaking and pronunciation), phonetics/phonology, and ESL pronunciation. Contact information: 4040 JFSB, BYU, Provo, UT 84602, 801.422.2938, Lynn_Henrichsen@byu.edu

Chirstin Stephens is a TESOL graduate student at Brigham Young University. She works at BYU’s English Language Center as a student mentor and part-time teacher. Contact information: 1715 E. 410 S. Spanish Fork, Utah, 84660, 801.798.7673, chirstin.s@gmail.com

REFERENCES

- de Bot, K., & Mailfert, K. (1982). The teaching of intonation: Fundamental research and classroom applications. *TESOL Quarterly*, 16, 71-77.
- Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition*, 20, 1-16.
- Derwing, T. M., & Munro, M. J. (2014). Windows of opportunity for L2 fluency and pronunciation development. Paper presented at the Pronunciation in Second Language Learning and Teaching (PSLLT) Conference, Santa Barbara, CA (September 6, 2014).
- Derwing, T. M., Munro, M. J., & Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning*, 48, 393-410.
- Derwing, T. M., & Rossiter, M. J. (2002). ESL learners’ perceptions of their pronunciation needs and strategies. *System*, 30, 155-166.

- Gilbert, J. (1987). Pronunciation and listening comprehension. In Morley, J. (Ed.), *Current perspectives on pronunciation: Practices anchored in theory* (pp. 29-39). Washington, DC: TESOL.
- Hahn, L. D. (2004). Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly*, 38, 201-223.
- Henrichsen, L. (2015). Video voiceovers for helpful, enjoyable pronunciation practice. In J. Levis, R. Mohammed, Z. Zhou, & M. Qian (Eds.), *Proceedings of the 6th Pronunciation in Second Language Learning and Teaching Conference*, University of California, Santa Barbara, Sept. 2014. (pp. 267-273), Ames, IA: Iowa State University.
- 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 (now available online at www.pronunciationmatters.com).
- Kang, O. (2010). ESL learners' attitudes toward pronunciation instruction and varieties of English. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 1st Pronunciation in Second Language Learning and Teaching Conference*, Iowa State University, Sept. 2009. (pp. 105-118), Ames, IA: Iowa State University.
- Levis, J., Link, S., & Sonsaat, S. (2013, October). Teaching pronunciation to advanced proficiency older learners: Building confidence and unlocking mysteries. Poster session presented at Second Language Research Forum (SLRF). Provo, UT.
- Macdonald, D., Yule, G., & Powers, M. (1994). Attempts to improve English L2 pronunciation: The variable effects of different types of instruction. *Language Learning*, 44, 75-100.
- Morley, J. (1991). The pronunciation component in teaching English to speakers of other languages. *TESOL Quarterly*, 25, 481-520.
- Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 45, 73-97.
- Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, 29, 191-215.
- Riney, T. J., Takagi, N., & Inutsuka, K. (2005). Phonetic parameters and perceptual judgments of accent in English by American and Japanese listeners. *TESOL Quarterly*, 39, 441-466.
- Purcell, E. T., & Suter, R. W. (1980). Predictors of pronunciation accuracy: A reexamination. *Language Learning*, 30, 271-287.
- Prator, C. H., & Robinett, B. W. (1985). *Manual of American English Pronunciation* (4th ed.). New York: Holt, Rinehart and Winston.
- Strauss, A. L., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Los Angeles: Sage.
- Suter, R. W. (1976). Predictors of pronunciation accuracy in second language learning. *Language Learning*, 26, 233-253.
- Yule, G., Hoffman, P., & Damico, J. (1987). Paying attention to pronunciation: The role of self-monitoring in perception. *TESOL Quarterly*, 21, 760-768.

Yule, G., & Macdonald, D. (1995). The different effects of pronunciation teaching. *International Review of Applied Linguistics*, 33, 345-350.

APPENDIX A

ESL 302, “Final Examination”

Instructions: Write a thoughtful response to each of the following questions. Do not worry about grammar, spelling, punctuation, etc. This is *not* a test of writing mechanics. For this examination, your *ideas, experiences, and feelings* are what count. Please respond freely and openly, as if you were talking to someone. As long as you answer honestly and adequately, your responses will be correct.

1. In what ways was this pronunciation class *valuable* to you? In what ways do you feel you have *improved* this semester because of this class?
2. Did this class *meet your expectations* of what a pronunciation class should be? Why or why not? What else would you have included in the class?
3. In what ways do you *practice* your English pronunciation (in speaking or listening mode) *outside* of ESL 302 class sessions?
4. Do you still use any of the *pronunciation-improvement strategies* we talked about earlier in the semester? Which ones? How often? When?
5. *How often* do you speak English with others (native English speakers and non-natives) each day? In what situations? Do you speak English more often now than you did before taking this class?
6. How *confident* are you in speaking English with native speakers? Why? What would help you gain more confidence?
7. What do you do in situations where you have a chance to speak with native English speakers and you are *uncertain they will understand* your pronunciation?
8. In what ways do you feel your *listening* has improved during the class? Do you hear English speakers more effectively now?
9. Describe a time recently when you were speaking with a native English speaker and experienced a *pronunciation challenge*. What happened? Was the communication successful? Why?
10. What things that you learned in this class this semester will you *continue to use in the future*?

Huang, M., & Pickering, L. (2015). Revisiting the pronunciation of English by speakers from Mainland China. In J. Levis, R. Mohammed, M. Qian & Z. Zhou (Eds). *Proceedings of the 6th Pronunciation in Second Language Learning and Teaching Conference* (ISSN 2380-9566), Santa Barbara, CA (pp. 206-216). Ames, IA: Iowa State University.

REVISITING THE PRONUNCIATION OF ENGLISH BY SPEAKERS FROM MAINLAND CHINA

[Meichan Huang](#), Texas A&M University – Commerce

[Lucy Pickering](#), Texas A&M University – Commerce

Pronunciation is a crucial component in achieving intelligibility and effective communication. Researchers have closely examined the pronunciation of English by Chinese speakers from various backgrounds, including Singapore, Hong Kong and Taiwan. However, less research has been conducted with Chinese speakers from Mainland China. The importance of studies of Mainland Chinese speakers cannot be underestimated as these speakers clearly make up the majority of ESL/EFL learners from this language background. This study builds on a previous study conducted with Mainland Chinese speakers. Deterding (2005, 2006) built a corpus of 19 Chinese speakers from major dialect areas of China, including northeastern provinces of Liaoning, Jilin; the eastern province of Shandong; and central provinces of Henan, Zhejiang, Jiangxi, Jiangsu, Anhui and Hunan. This study serves as a complementary study on the pronunciation characteristics of Chinese speakers from southern provinces in English. Three participants from Guangxi province were recorded reading a passage and participating in a short interview. Our goal is to compare our findings to those of Deterding (2006) and to suggest possible teaching applications for EFL in the southern provinces of China.

INTRODUCTION

Pronunciation features of Chinese English speakers from Hong Kong, Taiwan, and southern Asia have been well documented. Hung (2000) carefully studied segmental features of HK English, and confirmed that HK English has its own phonology in which the phonemic inventory is simpler than in inner circle varieties of English. Some research has also been done to document phonological features in connected speech. Deterding (2003) investigated the monophthong vowels of Singapore English using conversational vowels taken from ten Singapore English speakers, and found the distinction between [i] and [ɪ] was not maintained in Singapore English. Stibbard (2004) documented the co-occurring segmental errors, especially phonemic overlap in Hong Kong English. Apart from studies on Hong Kong English and Singapore English, Pennington & Ku (1993) examined the production of English final stops by Chinese speakers in Taiwan and concluded that the type of strategy used to produce English final stops varied according to the task type, place of articulation of final stop, age of subject, and native linguistic variety. Other studies have drawn attention to individual segmental features. For example, Wong and Setter (2002) acknowledged the possible conflation of [n] and [l] in syllable-initial position with HK English speakers. Peng and Setter (2000) analyzed in detail the alternation between the occurrence and absence of final consonants such as [t] and [d]. These studies have built up an overall image of English pronunciation of Chinese speakers, but it is yet to be completed by

adding features of Chinese speakers from Mainland China, which constitutes the largest proportion of Chinese speaking population.

Limited research has been conducted documenting English pronunciation of Mandarin speakers from Mainland China. In an early study, Chang (1987) listed features that Chinese speakers find problematic in their speech, such as vowels, consonants, clusters and intonation. However, this study did not include any recordings from Chinese speakers. Ho (2003) discussed pronunciation errors among PRC Chinese students and included students from many dialect areas, but his conclusions were based on indirect sources such as teacher observation and reflection. Hung (2005) confirmed in his preliminary study that [ʒ] was pronounced as [ɹ] by northern speakers. Qian (2011) conducted an acoustic investigation on segmental features of 12 students from northern China, identifying the absence of contrast between long and short vowels as a salient feature among the participants.

Deterding's (2005b, 2006) study on Chinese speakers is an exception, since he made the concept clear that there might be distinct differences in the English pronunciation produced by speakers of different dialects in China. Moreover, the research design captured pronunciation features in a naturalistic way. In his study, Deterding (2006) discusses extensively the English phonological features of 13 Chinese speakers from northeast, eastern, and central dialect areas of Mainland China. In the study, he concludes that there were twelve common features amongst the participants, ranging from segmental to suprasegmental features, such as extra final vowels and stressed final pronouns.

To better understand the English pronunciation of Chinese speakers from southern China, we undertook the following study. We hope that it will provide a better understanding of the English pronunciation of Mainland Chinese speakers from this area. Since Deterding's (2006) study has revealed much valuable information on PRC Chinese speakers' pronunciation features, we conducted a replication study amongst Chinese English speakers from the southern province of Guangxi. In this paper, we present some preliminary results that have distinguished southern Chinese English speakers from other speakers in China whose pronunciation features have been thoroughly discussed in Deterding's paper (2006).

METHODS

Participants

25 speakers from Guangxi, a southern province in China participated in this project (M=8, F=17). At the time they made the recordings, they were attending a one-month intensive English program in the same language institution at a local school.

In this paper, we present some preliminary results from three participants. The three participants were from three different cities (M=1, F=2). They were attending their undergraduate courses at different universities in Nanning, Guangxi at the time. The participant proficiency level was classified as intermediate (n=3) based on the fact that they had passed CET-4 exam (Note: the proficiency requirement to pass the national College English Test Level 4 is an intermediate level at minimum).

Two of the three participants reported speaking dialects of Guangxi at home, and all of them stated that they frequently used Mandarin Chinese in their school work. The dialects they used were classified under one main dialect area: Yue; which is under the same family as Cantonese.

According to the interview data and demographic questionnaire, two reported that they had never been abroad and they only use English occasionally or seldom. The other female student reported that she had been to Japan and used English during the trip.

Data Collection

Participants undertook two activities. The first activity was a passage titled “The Boy Who Cried Wolf”, which was suggested by Deterding (2006) (See Appendix A). The second part was a three-minute short interview. This interview was conducted to provide some additional spoken features that might not be presented in the Wolf passage. The topics of the interview, such as participants’ family and future career, were taken from the original recordings of Deterding’s (2005b) corpus. (See Appendix B).

The recordings were made in a quiet room, using an Olympus digital recorder VN-5200 PC with an attached HD microphone. This ensured a high recording quality, which enabled detailed phonetic and acoustic analysis of the data on Praat (Boersma & Weenink, 2005) and a Kay Pentax Computerized Speech Lab. The readings were coded phonetically by both authors to ensure the accuracy of the transcription. There was no disagreement between the two authors regarding the coding. The analysis focused on phonetic features previously identified by Deterding (2006) and any other features that were non-standard.

RESULTS

We found similar patterns in this study to those found in Deterding’s (2006) study. In Deterding’s paper, he identified a number of features in his participants’ speech that were different from native speakers of English, namely: extra final vowels, absence of reduced vowels, nasalized vowels, voiceless dental fricatives, voiced dental fricatives, the fricatives [v] and [z], vocalized [l], glide before [i], stress on function words and final pronouns, [h] pronounced as [x], [j] pronounced as [r], [l] and [n], and [d] or [z] as a replacement for voiced theta. We found a number of features in our three participants’ passage reading that are aligned with Deterding’s (2006) findings. In order for readers to compare our results, we have arranged our reporting in the same sequence as Deterding’s (2006) study. Following this, we report on a number of features that do not appear in Deterding’s data.

Findings Aligned with Deterding (2006)

Extra final vowels. An extra final vowel refers to “the addition of an extra vowel (an epenthetic vowel), usually a schwa, after a final plosive and before the next word” (Deterding, 2006, 179-180), therefore, *had* becomes [hædə]. This phenomenon was observed by Ho (2003) and extensively discussed in Deterding (2006). An extra final vowel was also noticed among the three participants, but it was not considered the most salient feature of their pronunciation. Although all three participants had the same problem, it was not distributed evenly. F2 and M3 had only 3 and 1 instance respectively whereas F1 had 16 instances. Two examples of F1’s speech are presented here:

1. ... a dark forest near the foot... (F1: 14.43s)

2. Raising his fist in the air...(F1: 25.32s)

When examining these examples, it became clear that all had a final plosive. In addition, 6 out of 16 of the instances with an extra final vowel ended with consonant clusters, such as *forest* (2 instances), *fist*, *feast*, and *convinced*.

Absence of Reduced Vowels

According to Deterding, reduced vowels (schwas) tend to occur in two contexts in British or American English. The first is “the unstressed syllables of polysyllabic words” (2006, p.182), such as *concern* in the Wolf passage. However, this passage only contains one instance of this. Thus, we set our eyes on the second context: “the weak force of monosyllabic function words” (ibid.), such as *that*, *than*, *to* and *of*. A total of 6 instances of *to*, and 6 instances of *of* were investigated for all the three participants. The realization of reduced vowels is as shown in Table 2, both F1 and F2 had a low rate of realization of reduced vowels on *to* (33%, 33%), and *of* (17% and 0%). M3 had an exceptionally high rate of reduced vowel production on *to* (83%), but an extremely low rate of realization of reduced vowel on *of*. These data show a clear trend of the participants producing full vowels instead of reduced vowels in function words.

Table 1

Realization of Vowel in “to” and “of”

speaker	to (6)		of (6)	
	Full	schwa	full	schwa
F1	4	2	5	1
F2	4	2	6	0
M3	1	5	6	0

Table 2

Nasalized Vowels in the Wolf Passage

speaker	Without nasalized [m]	Correct	Nasalized but not [m]
F1	0	2	2
F2	1	1	2
M3	0	3	1

Nasalized Vowels

Deterding (2006) noted a strong tendency of final vowels before final nasal consonants becoming nasalized or alternatively, a deletion of the final nasal consonant. Chung (2005) also discussed the tendency of English speakers in Taiwan for a tendency of deleting the final [n].

There are 4 instances of [m] in the Wolf passage, including 3 instances of *him* and 1 instance of *himself*. The full result for the 3 speakers is shown in Table 3. Of the three participants, F1 and F2 had a stronger tendency to nasalize the vowel, a 50% chance of nasalizing the vowel. Only F2 produced 1 instance with no nasal consonant [m]. Overall, only 50% of instances were produced correctly without any nasalization of final vowels or omission of nasal consonants.

Table 3

Realization of the Consonant at the Initial Voiceless Dental Fricative in the Wolf Passage

Instances	F1	F2	M3
thought	[s]	[s]	[s]
threaten	[s]	[s]	[tz]
Third	[s]	[θ]	[s]

Voiceless Dental Fricatives

There are three instances of [θ] in the passage, *thought*, *threaten*, and *third* respectively, making a total of 9 instances. The result shows that of the 9 tokens, 7 had clear instances of [s], 1 had [θ] and 1 had [tz]. The results of individual speakers are shown in Table 4.

Two of the speakers used [s] or [tz] with [θ] alternatively. Only F1 used [s] instead of [θ] throughout. Since there were no instances of final [θ] or middle [θ], we cannot conclude that the speakers would have the same tendency in each position. However, we can conclude is that the participants had a strong preference of pronouncing [s] instead of [θ].

Table 4

Realization of /ð/ at the Start of Words in the Wolf Passage

speaker	/ð/	/z/	/d/
F1	1	18	0
F2	0	0	19
M3	4	0	15

The results are similar to Deterding's (2006). Among a total of 57 tokens, only 5 tokens were pronounced with [ð]; either [d] and [z] has been used to substitute [ð]. Interestingly, although the three speakers are from the same province, they varied in their substitution of pronouncing [ð].

With regard to [ð] in middle and final word positions, all participants failed to produce [ð], instead, they tended to produce the initial and middle position [ð] with their expected pattern. However, only F1 and M3 pronounced with as [wIz] instead of [wIð], whereas F2 realized [ð] in the final position. It should also be noted that M3 did not follow the expected the pattern of producing [d] instead of [ð] in every word position. However, since there is only one representation of word final [ð], it is unclear if this is an anomaly or there is a different pattern in terms of how to pronounce word final [ð].

Voiced Dental Fricatives

There are 31 instances beginning with [ð] in the Wolf passage, 29 instances with initial [ð], 1 with middle [ð] (bother), and 1 with final [ð]. Here we considered the consonant at the start of 19 words: the (14 instances), that (3 instances), and they (2 instances). We chose these instances because each of them appeared at least twice in the reading. The result is as indicated in Table 5.

Table 5

Realization of Vocalized [ɪ] in 4 Instances of Wolf

speaker	Dark [ɪ]	Substitution of/no vocalized/ dark [ɪ]
F1	0	4
F2	0	4
M3	0	4

It is very clear that none of the participants realized the vocalized [ɪ] in the total 12 tokens. Thus, we can conclude that dark [ɪ] is a salient feature in the three participants' reading.

Vocalized [ɪ]

Deterding acknowledged that dark [ɪ] should not only be considered as a characteristic unique to Chinese speakers. Rather, vocalized [ɪ] is likely to become a trend in even standard English (Wells, 1982), just like the historical [ɪ] is no longer pronounced in words such as walk and calm.

In this passage, there are many instances of vocalized [ɪ]. For the sake of analysis, we chose 4 instances of *wolf* from the passage, and the result is shown in Table 6.

Table 6

Substitution of [w] with [v] in Village (2 tokens) and Villagers (2 tokens)

Speaker	Village (2)	Villagers (2)
F1	2	2
F2	1	2
M3	2	1

[w] as [v]

The last salient feature of the three students is [v] pronounced as [w]. Chang (1987), Hung (2005) and Deterding (2006) noted either replacement of [w] with [v], or omission of the consonant. In our data, a total of 12 tokens were taken from the readings. Only 17% of the tokens were pronounced as [v], and the majority of the tokens were replaced with either [w] or a very weak [v].

Listed above are the features that are aligned with Deterding's (2006) study on speakers from other parts of Mainland China. Since these features are not limited to southern speakers, we suggest that these features are shared among Mainland Chinese English speakers, regardless their varieties of Chinese.

New Features

In addition to the features discussed above, participants exhibited non-standard features that have not previously been reported. There are six categories of differences: Incorrect lexical stress, missing final stops, absence of distinction between long and short vowels, [ʃ] as [s], [v] as [f], and [s] as [k] in consonant cluster [ks]. In this section, we discuss these features in details.

Incorrect Lexical Stress

In the readings, two participants misplaced the stress in polysyllabic words. F1 and F2 misplaced the stress of *concern* on the first syllable. F1 also placed the stress of *actually* on the second syllable as indicated in the sentences below:

1. full of CONcern for his safety (F2: 52.09s)
2. It acTUally did come out of the forest (F1: 1:14.51s)

Although these are the only two instances examined in the reading, we will undertake further investigation in the interview data.

Omission of Final Stops

Two participants had a strong tendency to omit the final stops. In her reading, F2 omitted final stops in 15 instances, and 40% of the 15 instances comprised consonant clusters. M3 had 19 instances with this omission, and apart from instances that ended with consonant clusters (5

instances), the majority were monosyllabic words with a final stop. There was only one instance in which F1 omitted a final stop. We can infer from the data that there is a tendency for students from the southern provinces to omit the final stop in a word; however, this needs to be confirmed by further analysis with more participants.

Absence of the Distinction Between Long Vowels and Short Vowels

The absence of the distinction between long vowels and short vowels was recognized as a salient feature of pronunciation of Chinese speakers from the northern part of China (Qian, 2011). In our three participants, this is also a salient feature. For example, in their readings, the long vowel [i] was replaced with the short vowel [ɪ] in *feast* and vice versa in M3's reading.

Raising his fist in the air (F2: 36.58s) (M3: 34.20s)

And so the wolf had a feast (M3: 2:20.46s)

We also counted the number of instances each participant mixed long vowels with short vowels, and the number was quite noteworthy (F1: 5; F2: 9, M3: 5). We will undertake further analysis in the interview data.

[ʃ] as [s]

We found a total of three instances in which [ʃ] and [s] were used alternatively in F1 and M3's reading. For example, in this sentence:

...its fear of being shot, ... (M3: 39.28s)

shot was pronounced as [sɒt]. Meanwhile in F1's reading, *also* was pronounced as [ʌʃəʊ].

Although the number of instances is too scarce to make any significant conclusion, it remains likely that it is a pronunciation feature of English in southern province.

[v] as [f]

The final [v] in the preposition *of* was pronounced as [f] across the three participants in their readings. Out of the total 18 instances of the word *of*, all of them were pronounced with either clear [f] or a weak fricative [v]. Therefore, we believe that it is a common feature among the three participants, and it could possibly be a common feature among other participants.

[s] as [k]

In the readings, there was a tendency that the participants pronounce [k] in a consonant cluster of [ks] as [s] in the word *successful* (F2 and M3). This phenomenon could be interpreted in one of two ways. The first is that speakers substituted [k] with [s] in the consonant cluster [ks] because there is no consonant cluster in Mandarin Chinese and they could not pronounce it. The second is that they omit [k] for the same reason.

This second group of features are currently identified as unique to participants from the southern province and propose a direction for future analysis.

DISCUSSION

A comparison between the study reported here and Deterding's (2006) study on which this one is based, suggests that some pronunciation features are found to be shared among Chinese speakers across Mainland China. These features include extra final vowels, the absence of reduced vowels, nasalized vowels, voiceless dental fricative, vocalized [l], and [w] as [v]. Other features that were found in Deterding's (2006) study were not found thus far in this study. This may be due to the limited amount of data we have currently analyzed. Additional analysis is being undertaken with the interview data to determine if there are more features that our participants share with Chinese speakers from other parts of China. Finally, with regard to Deterding's findings, the choice of either [d] or [z] as a replacement of voiced dental fricative [ð] has been identified in the three participants' readings, which suggests that this feature may be pan-regional as opposed to region-dependent.

We also identified several features that are unique among our three participants from southern China. These features involve vowels, consonants, and stress placement, including omission of final stops, the absence of a distinction between long vowels and short vowels, misplaced stress, and three consonant replacements.

At this time, we have only included the analysis of the readings of three participants in this paper. Further analysis of the rest of the data will help us confirm some of our findings.

Concerning possible teaching applications based on our study, we would like to focus on certain segmental contrasts, such as voiceless/voiced dental fricatives and long/short vowels. We consider that using minimal pairs that contain problematic segmental contrasts will help Chinese students distinguish the individual confusing sounds. According to our data, some of the contrastive features that we can address using activities containing minimal pairs include: voiceless dental fricative [θ] with [s], voiceless dental fricatives [ð] with [d] or [z], long/short vowels, and [v] & [w]. By reading extensive word lists of minimal pairs that contain these problematic pairs of phonemes, and practicing with activities that are designed to drill the minimum pairs, we hope students can maximize their chance to produce these sounds correctly.

ABOUT THE AUTHORS

Meichan Huang is a second-year Ph.D. student in the Applied Linguistics Lab in Texas A&M University-Commerce. She is specializing in TESOL. Her interests include second language phonology with Chinese speakers and second language teaching, and semantics of humor.

Affiliation: Texas A&M University-Commerce

Address: Hall of Literature and Languages, P.O. Box 3011, Commerce, TX 75429-3011

Email: mhuang2@leomail.tamuc.edu

Lucy Pickering is an Associate Professor in TESOL and Applied Linguistics and director of the Applied Linguistics Laboratory at Texas A&M-Commerce. Her research program is focused on spoken discourse. She has done considerable work with Brazil's model of Discourse Intonation and its application to second language classroom discourse.

Affiliation: Texas A&M University-Commerce

Address: Hall of Literature and Languages, P.O. Box 3011, Commerce, TX 75429-3011
Email: Lucy.Pickering@tamuc.edu

REFERENCES

- Boersma, P., & Weenink, D. (2005). Praat. *Doing phonetics by computer*. [computer program]. Retrieved March, 2, 2015.
- Chang, J. (1987). Chinese speakers. In M. Swan & B. Smith (Eds.), *Learner English* (pp.224-237). Cambridge, England: Cambridge University Press.
- Deterding, D. (2003). An instrumental study of the monophthong vowels of Singapore English. *English World-Wide*, 24(1), 1-16.
- Deterding, D. (2005). Emergent patterns in the vowels of Singapore English. *English World-Wide*, 26(2), 179-197.
- Deterding, D. (2005b). A corpus of spoken PRC English. Retrieved from: <http://videoweb.nie.edu.sg/phonetic/prc-corpus/index.htm>.
- Deterding, D. (2006). The pronunciation of English by speakers from China. *English World-Wide*, 27(2), 175-198.
- Deterding, D. (2006). The North Wind versus a Wolf: short texts for the description and measurement of English pronunciation. *Journal of the International Phonetic Association*, 36(02), 187-196.
- Ho, L. (2003). Pronunciation problems of PRC students. In Lee et al. (Eds). *Teaching English to students from China*, 138-157.
- Hung, T. T. (2000). Towards a phonology of Hong Kong English. *World Englishes*, 19(3), 337-356.
- Hung, T. T. (2005). Word stress in Hong Kong English: A preliminary study. *HKBU Papers in Applied Language Studies*, 9, 29-40.
- Peng, L., & Setter, J. (2000). The emergence of systematicity in the English pronunciations of two Cantonese-speaking adults in Hong Kong. *English World-Wide*, 21(1), 81-108.
- Pennington, M. C., & Ku, P. Y. (1993). Realizations of English final stops by Chinese speakers in Taiwan. *RELJ Journal*, 24(2), 29-48.
- Qian, W. (2011). Phonological Features of China English: An Acoustic Investigation on Segmental Features of Educated China English Speakers. In *the Proceedings of The 16th Conference of Pan-Pacific Association of Applied Linguistics*.
- Stibbard, R. (2004). The spoken English of Hong Kong: A study of co-occurring segmental errors. *Language, Culture and Curriculum*, 17 (2), 127-142.
- Wells, J. C. (1982). *Accents of English* (Vol. 1). Cambridge: Cambridge University Press.
- Wong, C. S., & Setter, J. (2002). Is it 'night' or 'light'? How and why Cantonese speaking ESL learners confuse syllable-initial [n] and [l]. In *New Sounds 2000, subtitled Proceedings of the Fourth International Symposium on the Acquisition of Second Language Speech*. Klagenfurt: Universität Klagenfurt (pp. 351-359).

APPENDIX – A

The Boy who Cried Wolf

There was once a poor shepherd boy who used to watch his flocks in the fields next to a dark forest near the foot of a mountain. One hot afternoon, he thought up a good plan to get some company for himself and also have a little fun. Raising his fist in the air, he ran down to the village shouting ‘Wolf, Wolf.’ As soon as they heard him, the villagers all rushed from their homes, full of concern for his safety, and two of his cousins even stayed with him for a short while. This gave the boy so much pleasure that a few days later he tried exactly the same trick again, and once more he was successful. However, not long after, a wolf that had just escaped from the zoo was looking for a change from its usual diet of chicken and duck. So, overcoming its fear of being shot, it actually did come out from the forest and began to threaten the sheep. Racing down to the village, the boy of course cried out even louder than before. Unfortunately, as all the villagers were convinced that he was trying to fool them a third time, they told him, ‘Go away and don’t bother us again.’ And so the wolf had a feast.

APPENDIX – B

Informal interview script taken and modified from Deterding (2005b). The following are the range of topics that were covered in the interview:

1. Tell me something about your family.
2. What is your subject/major?
3. What do you want to do in the future?
4. What do you like to do in your free time?
5. Have you ever communicated with native speakers? Do you understand them? Do you think they understand you?

Miller, J., & Grim, F. (2015). Spoken French in a pronunciation course: Impressions and applications. In J. Levis, R. Mohammed, M. Qian & Z. Zhou (Eds). *Proceedings of the 6th Pronunciation in Second Language Learning and Teaching Conference* (ISSN 2380-9566), Santa Barbara, CA (pp. 217-228). Ames, IA: Iowa State University.

SPOKEN FRENCH IN A PRONUNCIATION COURSE: IMPRESSIONS AND APPLICATIONS

Jessica Miller, University of Wisconsin, Eau Claire
Frédérique Grim, Colorado State University

Standard French, also called international French, is typically taught in the second language (L2) classroom because native speakers of French from any geographical or social background should be able to understand it. Therefore, standard French gives learners a communication tool useful in most contexts. However, this arbitrary choice creates a linguistic dichotomy: native speakers, particularly those in the age group of our own L2 learners in a university setting, do not speak standard French. In this small-scale study, students enrolled in university French pronunciation courses share their opinions on standard and non-standard varieties of French in the second language classroom. Data were gathered through online pre-course and post-course questionnaires. A vast majority felt that being exposed to different varieties of French is important. In addition, most found standard and non-standard varieties equally important to learn, the former being a necessary foundation and the latter important building blocks giving relevance and authenticity to language learning. After discussing the results and pedagogical implications, activities on standard and non-standard language features are presented.

INTRODUCTION

A language is not, as we all know, a static and single entity. It is composed of many varieties. But it is neither logical nor practical to task second language (L2) students with learning them within the L2 curriculum. Instead, it is assumed that learners ought to be exposed to standard varieties. For example, standard French, also called international French, is typically taught because native speakers of French from any geographical or social background should be able to understand it. Therefore, standard French gives learners a communication tool useful in most contexts.

However, this arbitrary choice creates a linguistic dichotomy: native speakers, and in particular those in the age group of our own university L2 learners, do not speak standard French (Gadet & Guérin, 2008; Jeanmaire, 2014; Vitez, 2002). While they will understand our learners, will our learners understand them? In this small-scale study, students enrolled in university French pronunciation courses share their experience and opinion on standard and non-standard varieties of French in the second language classroom. Data were gathered through online pre-course and post-course questionnaires. Results are presented and pedagogical implications and applications for language courses are discussed.

Previous Research

Because language educators strive to train students to communicate effectively, the types of register and lexis taught are an important discussion point. As any language, French not only varies geographically (e.g., French from France, Switzerland, Québec, etc.), it also changes with

social influences (e.g., age groups, social class, gender, etc.). Which French should we teach and why? Few studies have looked into which variety of French is best suited within the communicative language approach to equip learners with tools that will increase their comprehension and intelligibility: should they be exposed to formal standard French or to spoken colloquial French? (Armstrong, 2001; Knaus & Nadasdi, 2001; Meissner, 1999).

Gadet and Guérin (2008) have referred to the oral mode as the “actualization of the language”, i.e., the language used authentically in a non-written, non-standard manner. They believe that both the written and oral modes have an important role in the classroom. Rather than opposing oral and written modes by associating them with phrases such as “non-standard/standard, incorrect/correct, deviant/normed, informal/formal” (p. 22, our translation), they described a continuum in which both are equally valued, but in different contexts. This sociolinguistic approach advises instructors to expose learners to many levels of the continuum, as they would be in authentic contexts. Teachers should also explain the values and implications of each level.

In terms of practical pedagogical implications, exposing students to various registers could be as simple as proposing alternate ways of speaking. Often we train students to view language use as right or wrong. That perspective may simplify language learning and teaching, but it does not portray languages and cultures as the rich entities that they truly are. In reality, language keeps changing, and what is considered wrong by some in one context on a given day, can be deemed right in other circumstances. Accepting that French is constantly changing should motivate teachers to include some new trends needed for authentic communication (Vitez, 2002).

Jeanmaire (2014) has recently questioned the normalization of the French language in regard to its actual development. He argues that because French cannot be constrained by rules and is in constant movement, French classes should reflect some of those current changes to equip learners with current authentic communicative tools. He particularly criticized calling language changes ‘mistakes’ and pointed to current rules that were initially considered barbarisms and which eventually became part of the norm. Jeanmaire suggested that learners of French should be exposed to language innovations such as borrowings or anglicisms as long as their usage is contextualized and meaningful.

For example, the French negative particle ‘ne’ has been disappearing in conversational French, making the secondary particle ‘pas’ a stronger marker of negation. This phenomenon is found among all age groups and social classes in France and Québec (Coveney, 1990). However, the vast majority of textbooks still teach the regular double negation *ne...pas* as the default negation. Occasionally, a note is added stating that ‘ne’ is dropped in oral French (Amon, Muyskens, & Hadley, 2015). But that fact is not emphasized in practice and learners may attribute too much importance to ‘ne’ and misunderstand oral French when it is absent. The question whether to actively teach our students to drop this particle in conversations was investigated by van Compernelle (2009). He remarked that discourse formality plays a strong role in the presence or omission of ‘ne’ in spoken French and suggests that teachers should prepare students to expect a variety of discourses to reach authentic communicative competence.

Research Questions

The discrepancies between ‘textbook French’ and ‘authentic French’ seem to widen as language change fails to be reflected in the classroom. Teaching standard French appears to be a safe choice: learners are potentially understood anywhere and, by using a neutral language in any

social context, they do not run the risk of using an inappropriate register that could embarrass them or others. Yet at the same time, they might feel excluded from native groups with whom they identify because they cannot interpret certain non-standard language cues, and therefore would be denied many important authentic learning experiences. The present investigation is thus motivated by the need to reconcile the practicality of teaching and learning a standard language with the authenticity of its use with native speakers. In order to find a path towards this reconciliation, we ask the following questions:

1. Do learners feel that being exposed to different varieties of French is important?
2. What variety of French is more important to learn in their opinion?

METHODS

Data were collected among adult students enrolled in three third-year semester-long pronunciation courses in three different American universities within their regular French language undergraduate programs. The content of those courses was similar as the three instructors, two of whom are the investigators of the present research, had collaborated to create teaching and learning material and were using the same textbook, *Sons et Sens: La Prononciation du Français en Contexte* (Violin-Wigent, Miller & Grim, 2013). They were free to use their preferred teaching strategies, but they all covered every phonetic and cultural theme of the textbook, thus exposing learners to both standard and non-standard varieties of French. Other than standard French, those varieties include familiar French (verlan, argot, slang, abbreviations, truncations), Québec French, Senegalese French, Meridional French, and Swiss French.

Questionnaires

Two online questionnaires were administered to gather data: one during the first week of the pronunciation class, and one, with adjusted related questions, during the last week of instruction. Responding to the questionnaires was strictly voluntary and done outside of class. As a result, end-of-semester attrition impacted our number of participants. Because many respondents who took the first questionnaire did not take the second one, the results should be taken with caution. However, all who took the second one also took the first one.

Participants

A total of 37 students completed to the pre-semester questionnaire, though only 13 of them returned for the post-semester follow-up questionnaire. The age of students who volunteered to provide data for this study ranged from 19 to 54 (median = 21). Most were native speakers of English, and none were native speakers of French.

RESULTS

At the beginning of the semester, participants answered the question “Do you think being exposed to different varieties of French is important?” A few students (8%) found it neither important nor important, while the rest (92%) thought it was rather or very important (Table 1).

One question that was repeated in both questionnaires was “In your opinion, what variety of French is more important to learn?” Before taking the pronunciation course (Table 2), a small majority (51%) favored standard French but others still believed that they were at least equally

important (49%). At the end of the term (Table 3), fewer favored standard French only (46%) and more thought that both were equally important (54%).

Table 4

Do you think being exposed to different varieties of French is important? (pre-course)

Response options	N	%
Not important at all	0	0%
Rather unimportant	0	0%
Neither important or unimportant	3	8%
Rather important	14	38%
Very important	20	54%

Table 5

In your opinion, what variety of French is more important to learn? (pre-course)

Response options	N	%
Standard French	19	51%
Regional varieties of French	0	0%
They are equally important	18	49%
Neither is important	0	0%

Table 6

In your opinion, what variety of French is more important to learn? (post-course)

Response options	N	%
Standard French	6	46%
Regional varieties of French	0	0%
They are equally important	7	54%
Neither is important	0	0%

In the second questionnaire our students were asked “Did you like being exposed to different varieties of French?” Over two thirds of them (69%) responded positively. The others expressed no preference (Table 4). In that same questionnaire we asked “If you have studied abroad, did

you sometimes have difficulties understanding or being understood because you were not familiar with conversational, familiar French?”. Only nine of them had studied in a French-speaking region before and were able to answer. Among them, a third (33%) said that they faced such a problem. Fewer (22%) were sure that they had not, while more (44%) were not sure (Table 5).

Table 7

Did you like being exposed to different varieties of French? (post-course)

Response options	N	%
Yes	9	69%
No	0	0%
No preference	4	31%

Table 5

If you have studied abroad, did you sometimes have difficulties understanding or being understood because you were not familiar with conversational, familiar French? (post-course)

Response options	N	%
Yes	3	33%
No	2	22%
Not sure	4	44%

The research questions detailed earlier can now be addressed. 1) Do learners feel that being exposed to different varieties of French is important? Yes, a vast majority (92%) thought so at the onset of their French pronunciation course. 2) What variety of French is more important to learn in their opinion? While most of them (51%) expressed a preference for standard French at the beginning of the course, this number decreased slightly (46%) after they had been exposed to different varieties and registers. Parallel to that, more students valued standard and non-standard as equals at the end of the semester than at the onset (up to 54% from 49%). Because there were fewer respondents in the follow-up survey, it is difficult to say if this trend would hold with a larger pool of participants.

DISCUSSION

As learners of French are rarely exposed to varieties other than standard French in the classroom it seems logical that our participants would favor learning standard French (51%), which is what they are familiar with. As they gained exposure and familiarity with other varieties throughout

the semester in the pronunciation class, so did they gain an appreciation for non-standard varieties. More students therefore attributed equal importance to those varieties at the end of the semester (54%) than at the beginning (49%).

A Basis as Building Blocks

Respondents' comments can help further interpret this finding. One student remarked: "Standard French is important to learn because it gives the basic building blocks to the language, but being able to take that knowledge and conform it to specific situations, i.e. with friends or family or at work, makes someone more capable at communication [...]." Another one wrote: "I think after taking this course it is important to grasp standard French with the regional varieties/informal French. I think standard French helps students grasp the main concepts and ideas, but that the regional varieties/informal French help to make the concepts relevant." Those observations underline the dichotomy between classroom vs. real life, or theory vs. practice. Both students describe the standard language as a basis on which to build, with the non-standard varieties serving as building blocks with which authentic French is constructed. In other words, standard French is seen as the necessary skeletal structure one needs to study in a classroom (much like a theoretical anatomy course) before being able to make sense of the true complexity of the language in authentic settings (like a medical practicum). A pronunciation course may help bridge those two stages by showing students dialectal varieties they had not noticed before, thus preparing them to expect language variation when they communicate with natives.

Another student's comment supports the idea that exposure to non-standard varieties prepares learners for more effective communication: "It's important to at least know that there will be variation between regions. And it will be helpful to know what to expect." Including different varieties of French in a third-year class is equipping learners with important linguistic and cultural tools that will help them be sensitive to variation when they encounter it, and perhaps ask questions about dialectal idiosyncrasies, thus enriching their learning experience and expanding their knowledge of the French language.

Helping Students Notice

In the end-of-semester questionnaire, a third of the respondents who had traveled in a French-speaking region reported having difficulties understanding or being understood because they were not familiar with conversational French. One of them notes: "I went to Paris and couldn't understand some of the conversational language because the person shortened the phrase and spoke quickly. Now I'm aware of liaisons, enchaînements, etc." This student seems to think that receiving training on some phonological features of the French language, specifically linking, could have increased his/her comprehension skills prior to traveling to France.

Indeed, many native English speakers perceive French as being spoken fast because syllable boundaries do not match word boundaries, as they do in English. As a consequence they listen for cues to parse words and interpret oral language that would work for English, but not for French. That technique can be successful with audio material targeting language learners, since it often provides slow and over-enunciated speech samples. However, the flow of an authentic conversation would throw off students unaccustomed to it and unfamiliar with its characteristics, and it does according to our data. Once learners are made to notice the right cues such as syllabic

structures and intonation patterns, comprehension of conversational French is facilitated. This type of parsing occurs in formal standard French, but has a complication in colloquial French in which optional schwas tend to be dropped, thus affecting syllabic structures. Without knowledge of how syllables function, it can be especially difficult for learners to interpret informal input.

Implications for Language Programs

Every student in this investigation believed that learning standard French is either more or equally important as learning non-standard varieties. Their comments, as explained above, support standard French as an essential foundation, and non-standard dialects as important building blocks. We can consequently recommend teaching standard French first without forgetting to include increased exposure to non-standard varieties in curricula.

At the same time, it is important to stress in which contexts those dialects or registers are appropriate as noted by Gadet and Guérin (2008). Without that crucial prior knowledge, it is not uncommon to have students returning from a semester abroad mixing formal and informal French in our classes. They pick up tidbits of conversational French without knowing how to categorize or handle them. For example, one of those returning students once described Voltaire as “un mec” (in English, a “dude”) in her/his French literature exam. That type of information would also benefit heritage speakers who are often comfortable with spoken French yet unable to switch to a more formal register for written projects. Even though students will likely not master registers by the end of a four-year university program, they should be made aware of their consequences. This information does not all have to be delivered and practiced in a pronunciation course. For a longer-lasting impact, it would be best to recycle it across various courses.

Further Research

Future studies should perhaps investigate what features of colloquial French have the most impact on comprehension and intelligibility, so that those characteristics can be taught in introductory courses. For example, negation is traditionally taught as requiring both ‘ne’ and ‘pas’ when in reality native French speakers often drop ‘ne’ and rely on ‘pas’ as the main cue to negation. Native English speakers tend to forget ‘pas’ and instead associate ‘ne’ with the negation as it resembles the English ‘not’ and is placed where ‘not’ or ‘don’t’ would be, which can create communication difficulties. When and how can negation be most effectively taught to avoid possible communication break-downs? Should other phonological rules concerning the schwa be taught (e.g., pronunciation of the masculine article ‘le’ and the preposition ‘de’)? In addition, teachers may wonder whether exposing learners to both standard and non-standard French might slow down their students’ movement up the ACTFL proficiency scale. As many institutions of higher education are concerned with preparing students to take an ACTFL Oral Proficiency Interview, one may object to exposing learners to more than one variety of French. On the other hand, a requirement of the Intermediate High and Advanced levels is for interviewees to be understood by non-sympathetic listeners. What phonological or syntactic features of colloquial French help non-sympathetic listener understand a non-native most easily?

Pronunciation Activities Promoting Exposure to Non-Standard Varieties

In a pronunciation class, certain phonetic themes lend themselves well to the exploration of non-standard varieties. For instance, learners can study syllabic structure (crucial in French to understand ‘enchaînement’, ‘liaison’, omission of schwas, and complementary mid-vowels) through activities around “verlan,” a way of creating words by inverting syllables in certain social contexts. The “verlan” dialect is itself in constant movement, and helping learners understand that the examples given in class may not be relevant when and where they travel is important to further demonstrate the constant evolution of a language.

The following activity (Table 6) is one way to bring to light current changes as recommended by Vitez (2002) and Jeanmaire (2014). This exercise introduces a non-standard variety of French, “verlan”, that students will encounter in authentic situations and can therefore be of value to them as they expressed in the present study. After students understand the preferred syllabic structures of French and are given a brief historical account of the development and formation of “verlan”, they can be given the following table to fill out. This task requires that they either derive the “verlan” form of a standard word or guess the original form of its “verlan” version. The top row provides a model. In the last row students can choose any word and produce its “verlan” counterpart.

Table 6

Verlan to teach French syllabic structures

Standard	Verlan
méchant	[ʃɑ̃.me]
	[si.mɛʁ]
maison	
comme ça	
tomber	

Understanding syllable structures in French plays an essential role in many innovations of conversational French. The next activity (Table 7) helps learners apply their knowledge of French syllables and the language preference for open syllables in order to predict the truncations of common words, in contrast to English that favors closed syllables. The last row is blank to allow students to come up with their own words.

Table 7

Truncations to teach French syllabic structures

Original	Truncated	English equivalent
réfrigérateur	frigo	
	labo	
vétérinaire		
adolescent		
interrogation		

Texting is a language that most college students know, use, and understand as an ever-changing code. This skill depends on understanding phoneme and grapheme correspondences. The next activity (Table 8) offers an opportunity for students to learn common texting codes while testing their understanding of the French sound system.

Table 8

Texting to teach French sounds

Standard	Text language
à demain	a2m1
	koi29
	Je tM
	TT où?
A+	
elle est au ciné	

The omission of the schwa in the negation particle ‘ne’, or even the omission of the entire word ‘ne’, needs to be taught in terms of type of discourse (van Compernelle, 2009). An activity offering sentences without ‘ne’ or with the absence of the schwa in ‘ne’ could ask students if they hear sentences in the positive or negative pole. Generally, the omission of schwa in any word is linked to informal contexts. The following listening activity (Table 9) helps learners associate the presence or absence of [ə] with the proper context. It teaches them to not expect to hear the letter <e> pronounced at all in everyday French. If they hear it, the situation is likely formal (e.g., a speech). If they don’t, the setting is likely informal (e.g., a conversation). This could also be done to distinguish standard French from Meridional French, since the latter tends to keep all schwas.

Table 9

Omitting schwa in informal contexts

Learner hears	Formal	Informal
Nous ne savons pas		
C'est cela		
Tout de même		
Il se trompe		
Donnez-moi le ticket		
J'ai repris l'argent		
Ils verront demain		
C'est la semaine prochaine		

In addition to using film to expose learners to various types of discourse as suggested by van Compernelle (2009), different genres of music can be proposed. Shadowing exercises can help pronunciation and expand vocabulary. For instance, a movie scene with the volume off can be played while learners reenact it live, trying to match the actors' lips as much as possible. Exercises as those presented here will draw attention to language variety, encourage students to create, and demonstrate that there is not necessarily a right or wrong answer, but rather a complex continuum of situations in which language types are more or less appropriate (Gadet & Guérin, 2008). Authentic resources are available to help, such as dictionaries that have recently given value to the language of the youth (Ribeiro 2014; Tengour, 2013).

CONCLUSION

Exposing second language learners to non-standard varieties supports the American Council on the Teaching of Foreign Languages (ACTFL) World-readiness standards for learning languages by giving learners tools for more effective Communication (standard #1), by representing the variety of Cultures (#2) of the Francophone world, by helping learners make Connections (#3) with other disciplines like geography and sociology, by helping them make Comparisons (#4) across cultures and linguistic varieties, and equipping them with language they can use to integrate real Communities (#5).

L2 Learners of French seem to value being exposed to non-standard varieties of French as long as their knowledge of standard French gives them a solid foundation. In this small-scale study no non-standard variety emerged as more important to explore than another. Students seemed to find value in being exposed to a wide array of dialects and registers. Deciding on which varieties to focus can be left up to the instructors or study-abroad programs, based on certain language features meaningful to course content or specific target regions. No matter what the students' goals in L2 language education, a clear map of linguistic diversity is essential for greater awareness and sensitivity of our global world.

ACKNOWLEDGMENTS

Jessica S. Miller received support from student differential tuition funds through the University of Wisconsin-Eau Claire's Faculty/Student Research Collaboration Differential Tuition Program.

We thank all students who were willing to participate in this study, as well as their professors who accommodated time to fill out the questionnaire.

ABOUT THE AUTHORS

Jessica Sertling Miller (Ph.D., University of Illinois at Urbana-Champaign), a native of Strasbourg, France, is an Associate Professor of French at the University of Wisconsin-Eau Claire. She teaches all levels of French undergraduate courses, with a focus on pronunciation, oral proficiency, and linguistics. She also supervises students in the French Education program. Contact Information: University of Wisconsin-Eau Claire, Department of Languages, Centennial Hall 4508, Eau Claire WI 54702-4004. Phone: (715) 254-1347. millerjs@uwec.edu

Frédérique Grim (Ph.D., University of Illinois at Urbana-Champaign) is an Associate Professor of French at Colorado State University. Her teaching foci are second language learning and teaching, French linguistics and phonetics. She advises the foreign language teacher program, the graduate program and co-directs the basic language program. Contact information: Colorado State University, Department of Foreign Languages and Literatures, C 104 Clark, Fort Collins CO 80523-1774. Phone: (970) 491-3867. Frederique.Grim@colostate.edu

REFERENCES

- Amon, E., Muyskens, J. A., & Hadley, A. C. O. (2015). *Vis-à-vis: Beginning French* (6th ed.). New York: McGraw-Hill Education.
- Armstrong, N. (2001). *Social and stylistic variation in spoken French: A comparative approach*. Amsterdam: John Benjamins.
- Coveney, A. (1990). The omission of *ne* in spoken French. *Francophonie*, 1, 2, 38-43.
- Gadet, F. & Guérin, E. (2008). Le couple oral / écrit dans une sociolinguistique à visée didactique. *Le français aujourd'hui*, 3, 162, 21-27.
- Jeanmaire, G. (2014). Le français en mouvement face à la norme. *The French Review*, 88, 1, 95-116.
- Knaus, V. and Nadasdi, T. (2001). Être ou ne pas être in immersion French. *The Canadian Modern Language Review*, 58, 2, 287-306.
- Meissner, F.-J. (1999). Vers l'intégration de la langue parlée dans l'enseignement du FLE. *Französisch heute*, 2, 155-165.
- Ribeiro, S. (2014). *Dictionnaire des ados français*. Paris : First éditions.
- Tengour, A. (2013). *Tout l'argot des banlieues: le dictionnaire de la zone en 2.600 définitions*. Paris : Les Éditions de l'Opportun.
- van Compernelle, R. (2009). Emphatic *ne* in informal spoken French and implications for foreign language pedagogy. *International Journal of Applied Linguistics*, 19, 1, 47-65.

- Violin-Wigent, A., Miller, J. & Grim, F. (2013). *Sons et Sens : Phonétique française en contexte*. Georgetown University Press.
- Vitez, P. (2002). The democratized norm of spoken French. *Linguistica*, 42, 1, 111-120.

INTEGRATION OF PRONUNCIATION IN FIRST-YEAR GERMAN TEXTBOOKS

[Iulia Pittman](#), Auburn University

This paper investigates the ways in which pronunciation is integrated into popular first-year German textbooks used in American colleges. Most students who register for a German college course have not had any experience with the German language. Research indicates that it is important to focus on pronunciation early in the learning process. First, college students of German are well past the critical age for language learning, and attaining good pronunciation in the foreign language becomes harder and harder with every passing year. Second, not focusing on pronunciation from the very beginning will allow fossilization to occur in students' speech. The longer these bad habits have to form, the harder it is to correct them later. A review of the ten most popular first-year German textbooks used in American colleges shows that only half of the textbooks include any information on pronunciation and only 20% of the textbooks present it in an effective way. The paper discusses possible reasons why this important aspect of language learning is not given more attention and offers recommendations for increasing awareness to the teaching of pronunciation in first-year German courses.

INTRODUCTION

According to the 2009 Modern Language Association (MLA) Enrollments in Languages other than English, German is the third-most commonly taught foreign language in American universities (with more than 96,000 students enrolled in a college German course). While a few programs use in-house materials, more than 97% of first-year German programs use one of the ten textbooks available on the market at the moment⁶. Because it is unreasonable (and not advisable) to expect a textbook to cover extensively every aspect of the language, the emphasis on different aspects of the German language may vary from textbook to textbook. Given that communication is a primary goal of language learning, we can expect to see considerable attention given to speaking. Thus, pronunciation, a subset of teaching students how to interact orally in the foreign language, should not be ignored (Henderson & Jarosz, 2014). The focus of this article is to explore the extent to which pronunciation is covered in German first-year textbooks and the ways in which it is integrated.

The importance of focused instruction

Research has found that explicit instruction is highly correlated with improved pronunciation in both a foreign and a second language (Saito, 2012). This improvement extends itself to comprehensibility, accent, and fluency (Derwing, Munro, & Wiebe, 1998). Kun-Ting et al. (2013) found that the technique of "shadowing" improved learners' pronunciation in English. In a study on form-focused instruction and corrective feedback on L2 pronunciation of the retroflex "r" by Japanese learners, Saito and Lyster (2012) found significant improvement in both

⁶ Source: anonymous editor from Vista Higher Learning

controlled speech and spontaneous speech. Saito's 2011 study showed significant improvement in the Japanese learners' comprehensibility as a result of explicit phonetic instruction. These findings corroborate Flege's (1999) strong recommendation to focus on pronunciation training in order to improve learners' pronunciation. However, in spite of all the findings that show why teaching pronunciation is important, in reality, it somehow "does not always make for a comfortable fit with instructors who support communicative language teaching" (Foote et al., 2013, 1). Among other factors, Henderson and Jarosz (2014) quote teachers' lack of training in teaching pronunciation. They further suggest two ways of assuring proper teaching of pronunciation to students. These are either providing a structured approach to pronunciation work in the textbooks or training teachers to better instruct their students.

Materials used in teaching pronunciation

The textbook used in a particular course can greatly affect the way a course is taught. This includes what is being covered, how it is being covered, and what and how material is being tested (Bragger & Rice, 2000). As mentioned in the above paragraph, teachers are not always comfortable teaching pronunciation, but well-written textbooks can help tremendously by guiding teachers in effective pronunciation teaching (Henderson & Jarosz, 2014). Some studies have been conducted to document the extent to which pronunciation is integrated in textbooks. One such study, on English as a Second Language (ESL), authored by Derwing et al. (2012) recommends teaching both segmental and suprasegmental features and focusing on sounds with a high functional load. The authors further suggest having a variety of tasks and including explicit explanations. Given the complexity of the process involved in modifying one's pronunciation, it is also recommended to repeat individual pronunciation features and to link them to other aspects of the language. Another study on first-year Spanish textbooks also stresses the importance of recycling phonetic topics and further recommends exposing students to different common Spanish dialects (Arteaga, 2000). Levac (1991) conducted a study in which she reviewed the integration of French vowels in four popular first-year textbooks. She found that the examples and explanations were often confusing and not offering students a good understanding of the pronunciation rules of French vowels. English as a foreign language in France and in Poland was the topic of investigation in an article authored by Henderson and Jarosz (2014). While suprasegmental features were well addressed, the authors found the overall integration of pronunciation unsatisfactory in the context of a communicative teaching method. They conclude their article with the observation that it is still up to the teacher to improve upon what the textbook lacks. This of course, brings us back to the challenge caused by "time-starved" teachers who feel ill equipped to design and deliver such work (276).

The importance of explicitly teaching pronunciation early

While there are still some unanswered questions with regard to the effect age has on successful foreign language learning, many studies have found support for the idea of multiple critical or sensitive periods (Huang 2014, Granena & Long 2013). Singleton and Lengyel (1995), for instance, suggest that there is no critical period for vocabulary. In other words, there is no point at which vocabulary acquisition stops. Bialystok and Hakuta (1994) state that certain syntactic features are affected by age, whereas others are not. Phonological acquisition is most affected by age (Oyama, 1976). Based on these observations, what is more relevant to the present study is the fact that the skill that is affected more strongly by age is pronunciation (Granena & Long, 2013). The majority of college students enrolled in German start the study of the language with

no prior knowledge and at a post-critical-period age⁷. While most significant differences are generally found between pre-adolescent and post-adolescent learners (Flege, 1987), attaining native-like pronunciation can become harder with every passing year for adult learners. Focused instruction on pronunciation at the beginning level can help in two ways. First, if all other things were equal, the effort it would take an 18 year old to reach a certain level of proficiency may be less than the effort it would take a 21 year old. Second, if pronunciation is first taught explicitly in the learners' third or fourth year⁸, learners will have had several years' time to possibly form some habits, often mispronunciations that are hard to break. The ingraining of these habits is not uncommon in second and foreign language learners' interlanguage speech and is known as fossilization, a concept introduced by Selinker (1972).

Unlike the development of morpho-syntactic features that often go through certain stages of development (for more information, see the Stages of Second Language Acquisition in Hill & Björk, 2008), when it comes to pronunciation, it is best if the learner is exposed to the new sounds and new suprasegmental features early on in the learning process (Arteaga, 2000; Dansereau, 1995). Research found that improved pronunciation helps with confidence in speaking and increased participation (Oyama, 1982a; Harlow & Muyskens, 1994), and it improves learners' comprehension (Meador, Flege, & MacKay, 1997). The earlier these aspects are affected positively, the better the learners can develop their foreign language skills.

A REVIEW OF THE TEN MOST POPULAR FIRST-YEAR GERMAN TEXTBOOKS

This overview includes the ten most popular textbooks used in American colleges to teach first-year German. This list includes:

1. *Sag mal*, Anton Ch. et al. (2014)
2. *Netzwerk*, Dengler et al. (2012)
3. *Treffpunkt Deutsch*, 6th edition, Gonglewski M. T. et al. (2012)
4. *Wie geht's*, 10th edition Sevin and Sevin (2010)
5. *Alles klar*, 2nd edition, Otto et al. (2003)
6. *Kontakte*, 7th edition Tschirner E. et al. (2013)
7. *Deutsch, na klar*, 6th edition, Di Donato, R. et al. (2011)
8. *Neue Horizonte*, 8th edition, Dollenmayer D. B. and Hansen Th. (2014)
9. *Vorsprung* 3rd edition, Lovik T. et al. (2014)
10. *Deutsch heute*, 10th edition, Moeller et al. (2012)

Criteria for analysis of the integration of pronunciation in the textbooks

Several criteria were selected to rate the selected textbooks on how they integrate pronunciation. While the absence of a pronunciation section in a textbook does not automatically mean that

⁷ It is important to note that a critical age period does not mean that children and adults past a certain age, usually puberty, cannot master a foreign language with native-like fluency. And the converse is also true: not all children exposed to a foreign language at an early age learn the language with native-like fluency.

⁸ In many German programs, focused courses on pronunciation are commonly taught in the third or fourth year of study.

pronunciation is not being addressed in the classroom, its presence increases the chances that it will be addressed. This is particularly the case given that the vast majority of textbooks have too much material, and instructors often feel pressured to finish what is included in the textbook, let alone add material that is not there. A first step of this overview, then, is to establish whether a textbook has a pronunciation section to begin with. The next step is to evaluate the *way* in which pronunciation is integrated into first-year German textbooks. A good presentation and integration is important for two reasons. First, learners have to be able to work with the material both inside and outside the classroom. Second, a good presentation of the material makes the teaching of this aspect of language considerably easier for instructors who both feel pressured for time and may not have much background in teaching pronunciation. The criteria selected for this analysis are: the existence of information on pronunciation at the chapter level, the presence of explanations, the presence of examples, the presence of modeling of pronunciation in an audio format available for students to listen to both in and outside of class, and the location and level of integration of the pronunciation section within the chapter. I will elaborate now on these points.

1. Information at the chapter level - In order to achieve any kind of success from teaching pronunciation in the first year, the information has to be present and embedded in every chapter. Having a few pages in an Appendix at the back of the textbook is not satisfactory. The information presented is to be broken up into manageable bites and addressed regularly. Ideally, both segmental and suprasegmental features are addressed.

2. Presence of explanations - German and English differ quite significantly in terms of their sound systems, and the correct pronunciation of German requires learners to produce sounds with new manners of articulation and in new places of articulation. Guidance that helps learning these new sounds can be helpful to the students in and outside of class.

3. Presence of examples - Talking about new sounds will only help to a certain extent. Like with other aspects of language learning, it is useful for learners to see the new sounds in context and receive comprehensible input. These examples ought to be words and sentences containing the newly introduced segmental and suprasegmental features. The context should be appropriate for the level of instruction and relevant to the chapter covered.

4. Modeling of pronunciation in an audio format - Correct modeling by the instructor in the classroom when presenting the new sounds is essential, and, without doubt, most instructors offer some sort of corrective feedback to their students regardless of the amount and way in which pronunciation is included in the textbook. However, students can benefit greatly from being able to listen to models of pronunciation outside of the class as well. This can be done by exposing students to audio recordings (on CDs or accessible online). In addition to focused exercises on certain segmental and suprasegmental features, audio recordings also have the benefit of allowing students to hear different native speaker accents of the target language.

5. Location and integration in the chapter - Where in the chapter the information on pronunciation occurs and how it is integrated can have a significant effect on whether it will be taught in class or not, and whether it helps students to learn the new sounds or not. It is important to present the new lesson on pronunciation relatively early in the chapter, to include vocabulary relevant to the chapter, and to make it an integral part of the chapter. Lessons of pronunciation are most effective when integrated into the chapters throughout the textbook, rather than added as a separate section at the end of the chapter (Derwing et al., 2012; Arteaga, 2000).

DATA COLLECTION

The collection of the data consisted of a review of all ten selected textbooks and an analysis of whether and how they were fulfilling the above-presented criteria. The data collection yielded the following results that are summarized in Table 1.

Table 1

Information on pronunciation found in all ten first-year German textbooks

Title	Each chapter	Explanations	Examples	Audio	Good location and integration	Other
1. <i>Sag mal</i>	Yes	Yes	Yes	Yes	Yes	-Suggestions -Opportunity to record -Regional variation
2. <i>Netzwerk</i>	Yes	Yes	Yes	Yes	Yes	-Deductive reasoning
3. <i>Treffpunkt Deutsch</i>	Yes	Yes	Yes	Yes	No	-Suggestions
4. <i>Wie geht's?</i>	Yes	No	Yes	Yes	Yes	-Suggestions
5. <i>Alles klar!</i>	9 out of 12	No	Yes	No	No	-Suggestions -Appendix section on pronunciation
6. <i>Kontakte</i>	No	-	-	-	-	- Appendix section on pronunciation
7. <i>Deutsch, Na klar!</i>	No	-	-	-	-	-
8. <i>Neue Horizonte</i>	No	-	-	-	-	-
9. <i>Vorsprung</i>	No	-	-	-	-	-
10. <i>Deutsch heute</i>	No	-	-	-	-	-

With each criterion being worth one point, each textbook was assigned a number of points depending on how many of the selected criteria it fulfills. A summary of the points earned by each textbook can be seen in Table 2.

Table 2

Textbook scores based on five criteria for integration of pronunciation

Textbook	Score
1. <i>Sag mal (SM)</i>	5
2. <i>Netzwerk (N)</i>	5
3. <i>Treffpunkt Deutsch (TD)</i>	4
4. <i>Wie geht's? (WG)</i>	4
5. <i>Alles klar! (AK)</i>	1.5
6. <i>Kontakte (K)</i>	0
7. <i>Deutsch, Na klar! (DNK)</i>	0
8. <i>Neue Horizonte (NH)</i>	0
9. <i>Vorsprung (V)</i>	0
10. <i>Deutsch heute (DH)</i>	0

RESULTS

When considering the five criteria selected for analysis, less than half (40%) of all textbooks meet at least four of the five criteria and only 20% of all books meet all criteria. The percentage of textbooks that have absolutely no pronunciation-related content at the chapter level is 50%⁹, while 60% of the textbooks have either no information, insufficient information, or information that is poorly presented and integrated.

⁹ In contrast, only 21.5% of ESL general skills textbooks do not contain sections on pronunciation (Foote et al., 2011).

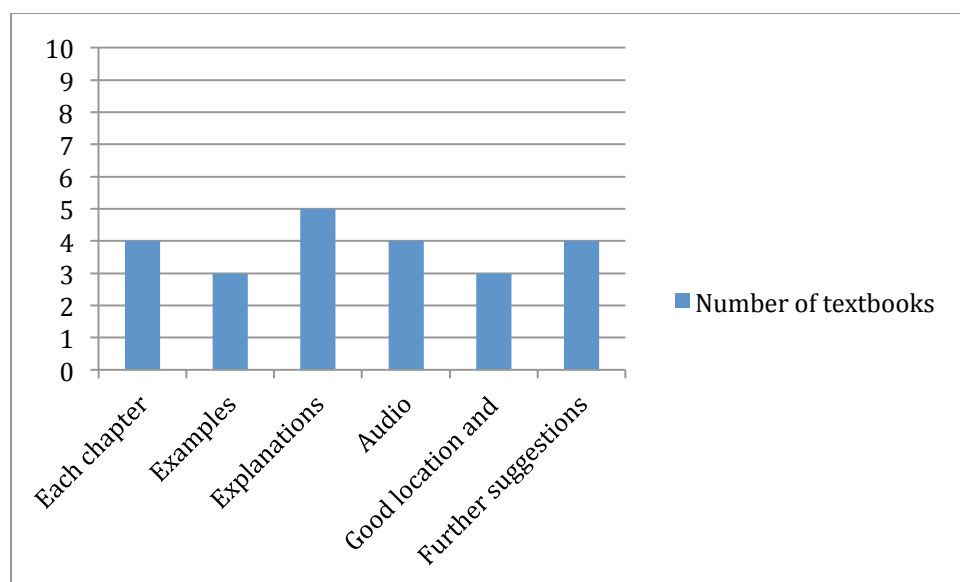


Figure 1. Number of textbooks fulfilling the selected criteria.

DISCUSSION

In this section, I will discuss the five criteria on which the textbooks were rated. The abbreviations for the textbooks are included in Table 2.

1. Information at the chapter level.

Four textbooks (SM, N, TD, and WG) have information on pronunciation in every chapter and a fifth textbook (AK) has such information in nine out of twelve chapters. The former four also clearly indicate the pronunciation sections labeled “Pronunciation” (*Aussprache*) in the table of contents, whereas AK simply lists the focus of the section (for example, “the ch sound” in chapter 2) which makes it a little harder to detect in the table of contents. This latter textbook also has a Pronunciation Guide in the Appendix. In contrast, four textbooks (DNK, NH, V, and DH) have no information on pronunciation whatsoever and one textbook (K) has an appendix at the end of the textbook.

2. Presence of explanations

Only three textbooks (SM, N, and TD) have explanations that accompany the new pronunciation lesson introduced. The explanations in SM und TD are in English, and they help students acquire new sounds in German or make subtle distinctions between English and German sounds. The third textbook, N, is published by a German publisher and is aimed at both non-native speakers of German in Germany and learners of German as a foreign language in many different countries. In order to accommodate many different native languages, the textbook is written completely in German. The extra guidance offered in this textbook is more deductive in nature. For instance, in the lesson on the distinction between short and long “e”, learners are to come up with the rule that applies to these sounds. The rule can be easily formulated with the help of a simple multiple-choice exercise that comes after an exercise in which students have the opportunity to listen to many examples of long and short “e” sounds. Knowing the time

constraints instructors face and lack of information on teaching how to articulate new sounds, the explanations found in these textbooks can make the teaching of pronunciation more effective.

3. Presence of examples

All five textbooks (SM, N, TD, WG, and AK) that include information at the chapter level about pronunciation also include examples with the specific segmental and suprasegmental features. The examples in SM consist of a variety of formats in every chapter. These include lists of words, sentences, and proverbs. In N, the format of the exercises varies from chapter to chapter. They include multiple choice exercises, lists of words, minimal pairs, sentences including the sounds under discussion, or even pictures to help elicit the right word choice from the student. TD includes generally one exercise with lists of words (which focuses on segmental features) and sometimes another exercise with sentences. The format in WG is fairly consistent throughout the textbook. The first 11 chapters include lists of words and minimal pairs, and the last four chapters only include lists of words. Similarly to TD, there are no suprasegmental features in this textbook. The examples (only segmental features) in AK are presented in a paragraph.

4. Opportunity to listen

Four (SM, N, TD, and WG) out of the five textbooks that have information at the chapter level also include audio recordings available for students. All four textbooks come with a CD so that students can listen to the different presentations of the new sounds. All recordings are read by native speakers. The textbook SM also offers students the opportunity to record themselves producing the new sounds through the accompanying Supersite and to submit the recordings to the instructor for feedback.

5. Location and integration

Of the five textbooks that include information on pronunciation at the chapter level, only three textbooks have a good location and integration of the information on pronunciation. SM consistently includes a well-written one-page section on pronunciation. These sections are integrated fairly early on in the chapter, and they have a predictable location and format. Although the pronunciation sections in N are significantly shorter, they also appear early in the chapter and use active vocabulary from the rest of the chapter. The pronunciation section in WG is located fairly early in the chapter, but often many of the words in the word lists or minimal pairs are irrelevant and unfamiliar to the students. The pronunciation section on TD is fairly long and well presented, but it is always at the end of the chapter. Not only do students not have the opportunity to practice the new sounds within the chapter, but given its location, one can see how an instructor pressured by time would find it easy to skip this section. While the pronunciation sections in AK are presented earlier in the chapter, they are not labeled well. They are labeled with “*Versuch’s mal*” (“You try it”), and given that there are no explanations, no CD exercises, no lists of words or minimal pairs to highlight certain sounds, it is hard to even distinguish that this is a pronunciation section.

Having said all the above, it is important to keep in mind that these findings tell us about what the textbooks include, but they do not necessarily give us an accurate picture of what is actually occurring in the first-year German classrooms. It is possible for an instructor to teach from a textbook that contains no information of pronunciation and yet bring excellent outside resources into the classroom. At the same time, it is possible for an instructor to teach from a textbook that covers pronunciation in an excellent way and for him or her to always skip these sections.

However, it is safe to assume that if the textbook integrates pronunciation well, the chances are higher that it will be taught to the students. The top-selling textbooks for first-year German on the Higher Education market are as follows:

Deutsch, na klar (ranked 7th from the top on my list of 10, with a score of 0 for pronunciation)

Kontakte (ranked 6th from the top, with a score of 0 for pronunciation)

Deutsch heute (ranked last, 10th out of 10, with a score of 0 for pronunciation)

Treffpunkt Deutsch (ranked 4th out of 10, with a score of 4 for pronunciation)

The results above show clearly that the most popular textbooks do not pay sufficient attention to integrating information on pronunciation. Overall, the findings corroborate findings of studies on textbooks for ESL, French, Spanish, and English as a foreign language in France and Poland, all of which show that pronunciation is an area that needs more focused attention in textbook authoring. In contrast with these studies, the current study found a greater prevalence of segmental features as the focus of pronunciation. In the two textbooks (SM and N) that include suprasegmental features, these constitute a small percentage of the lessons, and three textbooks (WG, TD, and AK) focus exclusively on segmental features.

Why is pronunciation not addressed to a greater extent in first-year German textbooks?

Textbook authors generally collaborate with the publisher during the writing process, but authors do have a great say in what goes into a textbook, considering that they are the experts in the field. One place to look for an answer for the variation in presence of pronunciation information in textbooks is the area of expertise of the authors of the textbooks themselves. An internet search was conducted to establish the authors' teaching and research expertise and to find out if their area of expertise had any correlation with the amount of pronunciation integrated into the various textbooks. Information was gathered from university bio pages and the textbook bio information, when available. The following table summarizes the areas of expertise of each author for each of the ten selected textbooks.

Table 3

Textbook authors and their teaching/research background

	Literature	SLA	Pedagogy	Other
1. Sag mal , Author 1	Yes		Yes	
1. Sag mal , Author 2		Yes		
1. Sag mal , Author 3		Yes		
1. Sag mal , Author 4		Yes		
2. Netzwerk , Author 1		Yes		
2. Netzwerk , Author 2		Yes		
2. Netzwerk , Author 3		Yes		

2. Netzwerk , Author 4		Yes		
3. Treffpunkt Deutsch , Author 1		Yes		
3. Treffpunkt Deutsch , Author 2		Yes		
3. Treffpunkt Deutsch , Author 3	Yes (culture)			
4. Wie geht's , Author 1	Yes		Yes	
4. Wie geht's , Author 2				Yes (conversation)
5. Alles klar , Author 1	Yes		Yes	
5. Alles klar , Author 2	Yes			
5. Alles klar , Author 3	Yes		Yes	
5. Alles klar , Author 4			Yes	
6. Deutsch, na klar , Author 1			Yes	
6. Deutsch, na klar , Author 2	Yes		Yes	
6. Deutsch, na klar , Author 3	Yes		Yes	
7. Kontakte , Author 1		Yes		
7. Kontakte , Author 2	?	?	?	?
7. Kontakte , Author 3		Yes		
8. Neue Horizonte , Author 1	Yes			
8. Neue Horizonte , Author 2	Yes			
9. Vorsprung , Author 1		Yes		
9. Vorsprung , Author 2				Yes (photography)
9. Vorsprung , Author 3		Yes		
10. Deutsch heute , Author 1			Yes	
10. Deutsch heute , Author 2		Yes		
10. Deutsch heute , Author 3	Yes			
10. Deutsch heute , Author 4				Yes (Authentic materials)
10. Deutsch heute , Author 5				Yes (creative activities)
Total	11	14	9	4
Only one	5	14	3	4

Out of the 33 authors, five authors (15%) name literature as their main and only area of expertise, fourteen (42%) name second language acquisition or linguistics as their main area of expertise, three (9%) consider pedagogy, and five authors (15%) list other areas. It is common for professors to have teaching and research interests in more than one field. As such, the table shows that six authors (18%) do work in both the fields of literature and pedagogy. The percentage of authors with various backgrounds was compared to the scores the textbooks received for pronunciation. It must be noted that, since it is not always clear what the background of the authors in the “Other field” column is, and there is variation within the group, these authors were left out of this analysis. The textbooks were divided into two groups, “high” (the four textbooks that received a score of higher than 4) and “low” (the six textbooks that received a score lower than 2). It was found that author background affects the degree to which pronunciation is integrated into textbooks. Specifically, the chances were significantly higher (75%) for a textbook to include information on pronunciation if the author’s background was second language acquisition or linguistics. In contrast, the percentages of authors of textbooks that scored high on pronunciation were much lower for the areas of literature (8%), pedagogy (0%), or both (17%). The percentages of authors in the “low” group did not vary much one from the other in terms of author’s area of expertise.

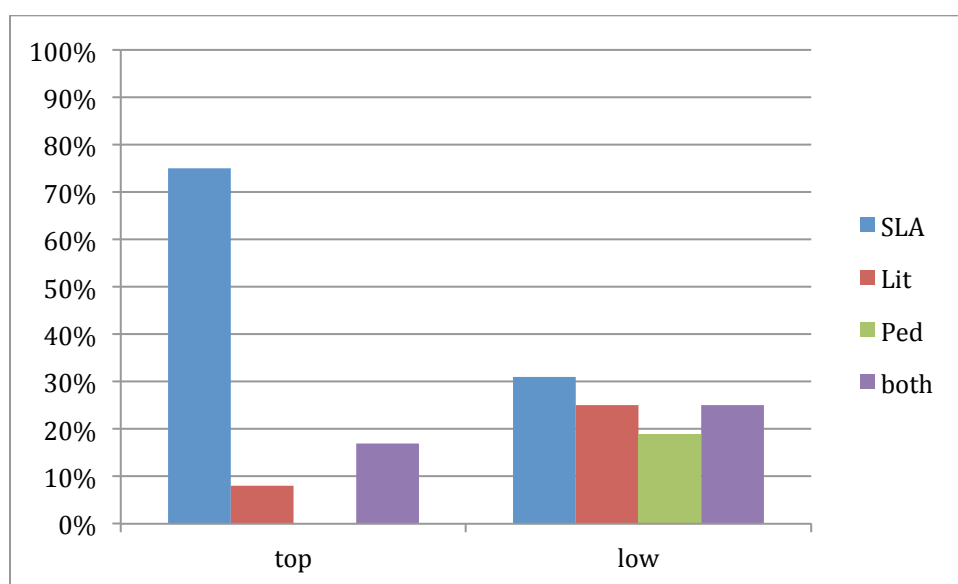


Figure 2. Percentage of authors with various backgrounds in “high” and “low” textbook category.

Implications of lack of information on pronunciation in first-year German textbooks

As mentioned above, given the fact that most textbooks are already overloaded with information, unless it is on the instructors’ radar to begin with, chances are high that they will not add sections on pronunciation to their busy curriculum if it is not already included in the textbook.

Recommendations

The first recommendation is to raise publishers' and textbook authors' awareness of the importance of integrating information on pronunciation into future editions. Second, when selecting new books, faculty members should pay attention to this aspect of foreign language teaching and select textbooks that have a good section on pronunciation. Third, when working with an otherwise great textbook that has comparatively little pronunciation information, instructors should strongly consider supplementing with outside information.

Future study

I would like to continue this study with an in-depth investigation of what is happening in the classroom. Information from instructors from a large number of institutions collected via electronic survey can supply insightful information about how pronunciation is taught in class. The data can further tell us how much of what is going on in the classroom depends on the textbook used.

CONCLUSION

The findings of this paper show that surprisingly little attention is given to the integration of pronunciation in first-year German textbooks. Lack of addressing focused pronunciation at the elementary level or delaying it to the advanced levels can result in students experiencing fossilization of their pronunciation mistakes. Instructors are recommended to use supplementary materials whenever there is a need.

ABOUT THE AUTHOR

Iulia Pittman (pittman@auburn.edu) is an Associate Professor of German and Linguistics at Auburn University. A native of Romania, she grew up bilingually speaking Romanian and Hungarian. Her research interests include bilingualism, code-switching, foreign accents, and technology in foreign language teaching. She is the author of the book *The Romanian Accent in English*, the editor of and contributor to the volume *Raising Children Bilingually in the United States*, author of a journal article on trilingual code-switching that appeared in the *International Journal of Multilingualism*, and author of several other articles.

REFERENCES

- Arteaga, D. L. (2000). Articulatory phonetics in the first-year Spanish classroom. *Modern Language Journal*, 84, 339-354.
- Bialystok, E. and Hakuta, K. (1994). *In other words: The science and psychology of second language acquisition*. New York: Basic Books.
- Bragger, J.D., and Rice, D.B. (2000). Foreign language materials: Yesterday, today, and tomorrow. In R.M. Terry (Ed.), *Agents of change in a changing age* (pp. 107-140). Lincolnwood, IL: National Textbook Company.
- Dansereau, D. (1995). Phonetics in the beginning and intermediate oral proficiency-oriented French classroom. *French Review*, 68, 638-651.

- Derwing, T. M., Diepenbroek, L. G., & Foote, J. A. (2012). How well do general-skills ESL textbooks address pronunciation? *TESL Canada Journal*, 30(1), 22–44.
- Derwing, T. M., Munro, M. J., & Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning*, 48(3), 393–410.
- Flege, J. (1999). Age of learning and second-language speech. In D. Birdsong (Ed.), *Second language acquisition and the critical period hypothesis* (pp. 101–132). Hillsdale, NJ: Lawrence Erlbaum.
- Flege, J. E. (1987). A critical period for learning to pronounce foreign languages? *Applied Linguistics*, 8, 162–177.
- Foote, J. A., Trofimovich, P., Collins, L., & F.S. Urzúa (2013). Pronunciation teaching practices in communicative second language classes. *The Language Learning Journal*, 1–16.
- Foote, J.A., Holtby, A.K., & Derwing, T.M. (2011b). Survey of the teaching of pronunciation in adult ESL programs in Canada, 2010. *TESL Canada Journal*, 29, 1–22.
- Furman, N., Goldberg, D., & Lusin, N. (2009). Enrollments in languages other than English in United States institutions of higher education. *Modern Language Association*. MLA, Dec. 2010. Web. 3 Jan. 2011.
- Granena, G., & Long, M. (2013). Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research*, 29(3), 311–343.
- Harlow, L. L., & Muyskens, J. A. (1994). Priorities for intermediate-level language instruction. *Modern Language Journal*, 78(2), 141–154.
- Henderson, A., & Jarosz, A. (2014). Desperately seeking a communicative approach: English pronunciation in a sample of French and Polish secondary school textbooks. In E. Waniek-Klimczak, (Ed.), *Research in Language*, 12, 261–278.
- Hill, J. D., & Björk, C. L. (2008). *Classroom instruction that works with English language learners: Facilitator's guide*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Huang, B. (2014). The effects of age on second language grammar and speech production. *Journal of Psycholinguistic Research*, 43(4), 397–420.
- Kun-Ting, H., Da-Hui, D., & Li-Yi, W. (2013). A preliminary study of applying shadowing technique to English intonation instruction. *Taiwan Journal of Linguistics*, 11(2), 43–65.
- Levac, P. (1991). Teaching languages communicatively: An examination of the presentation of phonology to first-year students of French. In J. Alatis (Ed.), *Georgetown University Round Table on Languages and Linguistics* (pp. 215–222). Washington, DC: Georgetown University Press.
- Oyama, S. (1982a). A sensitive period for the acquisition of a nonnative phonological system. In S. Krashen, R. Scarcella, & M. Long (Eds.), *Child-adult differences in second language acquisition* (pp. 20–38). Rowley, MA: Newbury House.
- Oyama, S. (1976). A sensitive period for the acquisition of a nonnative phonological system. *Journal of Psycholinguistics Research*, 5, 261–283.

- Meador, D., Flege, J. E., & MacKay, I. R. A. (1997). Nonnatives' perception of English sentences presented in noise. *Journal of the Acoustical Society of America*, 101(5), 3129-3130.
- Saito, K. (2012). Effects of instructions on L2 pronunciation development: A synthesis of 15 quasi-experimental intervention studies. *TESOL Quarterly* 46(4), 824-854.
- Saito, K., & Lyster, R. (2012). Effects of form-focused instruction and corrective feedback on L2 pronunciation development of /ɹ/ by Japanese learners of English. *Language Learning*, 62(2), 595-633.
- 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.
- Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics*, 10, 209-231.
- Singleton, D., & Lengyel, Zs. (1995). *The age factor in second language acquisition*. Clevedon. Multilingual Matters.

TEXTBOOKS

1. **Sag mal**, Anton Ch. et al. (2014) Vista Higher Learning, Boston, MA
2. **Netzwerk**, Dengler et al. (2012) Langenscheidt, Berlin und München, Deutschland
3. **Treffpunkt Deutsch**, 6th edition, Gonglewski M. T. et al. (2012) Prentice Hall, Upper Saddle River, NJ
4. **Wie geht's**, 10th edition Sevin and Sevin (2010), Cengage Learning, Boston, MA
5. **Alles klar**, 2nd edition, Otto et al. (2003), Pearson/Prentice Hall, Upper Saddle River, NJ
6. **Kontakte**, 7th edition Tschirner E. et al. (2013), McGraw Hill. New York, NY
7. **Deutsch, na klar**, 6th edition, Di Donato, R. et al. (2011), McGraw Hill, New York, NY
8. **Neue Horizonte**, 8th edition, Dollenmayer D. B. and Hansen Th. (2014) Cengage Learning, Boston, MA
9. **Vorsprung** 3rd edition, Lovik T. et al. (2014) Heinle Cengage Learning, Boston, MA
10. **Deutsch heute**, 10th edition, Moeller et al. (2012) Cengage Learning, Boston, MA

Thir, V. (2015). Implications of English as a Lingua Franca for pronunciation teaching in teacher education. In J. Levis, R. Mohammed, M. Qian & Z. Zhou (Eds). *Proceedings of the 6th Pronunciation in Second Language Learning and Teaching Conference* (ISSN 2380-9566), Santa Barbara, CA (pp. 243-252). Ames, IA: Iowa State University.

IMPLICATIONS OF ENGLISH AS A LINGUA FRANCA FOR PRONUNCIATION TEACHING IN TEACHER EDUCATION

Veronika Thir, University of Vienna

With the majority of conversations in English these days taking place in international settings, TESOL professionals have increasingly come to realize the importance of English as a Lingua Franca (ELF) communication for learners of English. Yet, it seems that the practices of English pronunciation teaching have still remained largely unaffected by these developments, with NS models prevailing in the majority of ESL/EFL teaching contexts while the implications of ELF for pronunciation often remain unconsidered. This paper suggests how pronunciation teaching in non-native speaker teacher education could be updated to equip future NNS teachers with the knowledge and skills necessary to make informed decisions for English pronunciation teaching in a globalized world. It is argued that this can best be achieved by a combination of pronunciation training, theoretical education, and critical reflection. Furthermore, teacher education should help NNS teachers to develop a positive professional identity as English pronunciation teachers, which might be achieved by educating NNS teachers about their status as legitimate NNS pronunciation models and their potential to teach English pronunciation effectively without speaking with a ‘native-like’ accent.

INTRODUCTION

The global spread of English during the past decades has led to an immense increase of non-native learners of English around the world. The large majority of these learners use English primarily not as a second or foreign language in communication with native speakers (NSs) of the language, but as a *lingua franca* mainly with other non-native speakers (NNSs) of various cultural and linguistic backgrounds. This type of English is nowadays commonly referred to as English as a lingua franca (ELF). Although ELF research is still a comparably young field, scientific interest in ELF has increased dramatically during the past decade (Jenkins, 2009, p. 143). This might be explained by a growing awareness of ELF as a phenomenon that concerns speakers of English in all three of Kachru’s ‘Circles’ (1985), i.e. NNSs in the so-called ‘Expanding Circle’ and the ‘Outer Circle’ (the former referring to countries where English has traditionally been learned as a ‘foreign’ language, the latter to countries where English is an institutionalized ‘second’ language, often having the status of an official language), but also NSs from ‘Inner Circle’ countries, where English is spoken as a first language (such as the US or the UK). Being restricted neither to a particular social context nor to a specific geographical region, ELF communication takes place all around the globe, wherever speakers of English of various different L1 backgrounds meet and communicate with each other. In 2001, Seidlhofer considered it “the most extensive contemporary use of English worldwide” (p. 133), and with the continuing increase of globalization and the consequent need for a common medium of international communication, ELF will most probably continue to expand further in the future. Obviously, this has important implications for English language teaching (ELT), as numerous learners are now

to be prepared for ELF communication rather than for ‘foreign language’ communication with NSs of English.

Pronunciation and ELF

If learners of English are to be prepared adequately for ELF interactions, particular attention is to be accorded to pronunciation teaching. Empirical research by Jenkins (2000) found that pronunciation was the area of language most crucial to successful ELF communication: over two thirds (27 out of 40) of communication breakdowns in Jenkins’ data were due to pronunciation errors. Research by Deterding (2013) on South-East Asian ELF confirmed this tendency, with 86 % of misunderstandings in his data involving pronunciation errors.

Based on her findings, Jenkins (2000) devised a pedagogical ‘core’ of sounds essential to intelligibility in ELF. This ‘Lingua Franca Core’ (LFC) includes (cf. Jenkins, 2000, p. 159):

- all consonant sounds, except for /θ/, /ð/ and [ʃ]
- vowel length contrasts (including those caused by a following lenis or fortis consonant)
- aspiration of /p/, /t/, and /k/
- the long central vowel /ɜ:/
- maintaining initial consonant clusters (word-medial or final consonant clusters may only be simplified according to L1 rules of elision)
- rhoticity and avoidance of [r] for /r/ in intervocalic position
- nuclear stress placement and chunking

As can be seen from the above list, it is the segmental rather than the suprasegmental level that was found to be crucial to intelligibility in ELF communication. Indeed, virtually all instances of communication breakdown in Jenkins’ data involving pronunciation were caused by negative L1 transfer on the segmental level (Jenkins, 2000, pp. 87-88). Jenkins attributes the importance of the segmental level for intelligibility in ELF to the fact that NNSs tend to rely on bottom-up rather than top-down processing strategies, “which, in turn, lead them to focus too firmly on the acoustic signal” (Jenkins, 2000, p. 20).

It should be noted that the above list is not to be regarded as definite, and the need for further empirical research to confirm and ‘fine-tune’ the LFC has been emphasized by Jenkins herself (2000, p. 235). A number of studies have already been conducted for this purpose: for example, the findings of Deterding (2013), Deterding and Kirkpatrick (2006), Osimk (2009), and Rajadurai (2006) all confirmed that /θ/ and /ð/ are not essential for intelligibility in ELF communication. The importance of aspiration of /p/, /t/, /k/ was confirmed by Osimk’s and Rajadurai’s research, but could not be justified on the basis of Deterding’s (2013) data. In this sense, the LFC is to be understood as an “ongoing empirical description of how non-native speakers achieve mutual intelligibility” (Walker, 2010, p. 44).

Another important aspect of Jenkins’ proposal for pronunciation teaching is her call for the teaching of phonological accommodation skills, which, according to her data, are essential for successful ELF communication. However, despite their communicative value, such phonological

adjustments are only rarely used by ELF speakers (Jenkins, 2000, p. 180-181). Jenkins therefore stresses the need to teach both productive and receptive phonological accommodation skills to learners who wish to engage in international communication (Jenkins, 2000, p. 210, 2005, p. 150; see Walker, 2010, for practical teaching suggestions). However, it seems that this aspect of Jenkins' proposal is frequently overlooked, in particular by those who mistake the LFC for a strict and invariable pronunciation norm to which learners should conform regardless of the particular situational context in which they find themselves and the type of interlocutors with whom they are communicating. It should be noted that this idea of the LFC would go against the very concept of ELF itself, as linguistic norms in ELF are not determined by external (NS or NNS) norms, but negotiated by the interlocutors themselves, being "established during the interaction within the current possibilities"¹⁰ (Seidlhofer, 2011, p. 18). Jenkins herself affirmed from the beginning that she did not intend the LFC to be taken as a model for imitation, but as a pedagogical core approach that allows for variety in L2 pronunciation (Jenkins, 2000, p. 131; see also Jenkins, 2005, p. 147 & p. 151, 2006, p. 36). After all, one of the most important intentions behind Jenkins' LFC was to give NNSs "the same sociolinguistic rights as those enjoyed by L1 speakers" (Jenkins, 2005, p. 147) by allowing them to express their L1 identity via their accent. This is possible via the 'non-core features' (i.e. those pronunciation features not included in the LFC, such as vowel quality or the th-sounds), which are intended to remain open to a speaker's personal preference.

Implications for English Language Teaching

The rise of ELF has serious implications for the teaching and learning of English in general and English pronunciation in particular. Within traditional approaches to English language teaching, it has been assumed that NNS learners of English are to be prepared for communication with NSs. Yet, such a perspective seems limited in the light of the fact that most NNS learners nowadays primarily communicate with other NNSs of English. In other words, the question arises why English pronunciation teaching should persist in exclusively orienting its goals and models towards NS usage when numerous NNSs successfully communicate with each other every day without adhering to NS pronunciation norms. This is not to say that NS norms have lost their validity for ELT altogether, for there will always be learners of English whose primary interest lies in communicating with NSs and for whom a NS accent will hence constitute the most appropriate pronunciation model. Rather, it means that for learners who are more likely to use English as a *lingua franca* to interact (mostly) with other NNSs rather than as a foreign language (and this will be the vast majority of NNS learners of English), a different approach to teaching English pronunciation might have to be considered, namely one that privileges (or at least gives equal importance to) the implications of ELF for users of English instead of focusing solely on the teaching of the linguistic norms of ENL varieties. For this to happen, ELF and its implications for ELT will have to form part of teacher education and teacher training programs, so that future teachers of English will be able to make informed decisions as to the teaching

¹⁰ This means that rather than strictly conforming to NS usage, successful ELF speakers settle on 'ad hoc' norms that are operable for all participants involved and that depend on the common linguistic resources available to them (see further Seidlhofer, 2011, p. 18). On the level of pronunciation, this kind of negotiation of linguistic norms can best be described as the aforementioned 'phonological accommodation', which involves productive phonological accommodation (the readiness and ability to adjust one's pronunciation to the perceptive needs of one's interlocutor) and receptive phonological accommodation (the readiness and ability to accommodate receptively to the idiosyncrasies of one's interlocutor's pronunciation).

goals and pronunciation models appropriate in a specific teaching context (Jenkins, 1998b, 2000). Yet, in numerous teacher education programs aimed at NNSs of English, there has been a strong focus on pronunciation training rather than teacher education, with socio-psychological and socio-linguistic considerations only assuming a marginal role or being neglected altogether. At the Vienna English Department, for example, teaching degree students receive extensive phonetic training in either Received Pronunciation (RP) or General American (GA), with the course aim being a pronunciation approximating the chosen model accent as closely as possible. The general applicability of both RP and GA as models in all teaching contexts remains unquestioned throughout the course, and alternative approaches to English pronunciation teaching or the downsides of the so-called ‘nativeness principle’ (Levis, 2005) remain unconsidered (cf. Thir, 2014). Such normative approaches to English pronunciation teaching do not only seem outmoded in the light of the unprecedented sociolinguistic developments that have taken place in the English speaking world in the past decades, but potentially detrimental to the development of NNS teachers’ professional identity, who, unless they manage to attain a native-like accent in English (an ambitious goal achieved by few), often seem to feel concerned about their own pronunciation skills (Medgyes, 1994; Tang, 1997; Kamhi-Stein, Aagard, Ching, Paik, & Sasser, 2004), hence doubting their competence to teach English pronunciation effectively (. Yet, the active involvement of NNS teachers in English pronunciation teaching is crucial for the success of the ELT enterprise, as the majority of foreign learners of English are taught by NNSs of English (Miller, 2009, p. 176; Moussu & Llurda, 2008, p. 315), and these learners will not receive the necessary instruction to develop the pronunciation skills needed in a globalized world (and/or in a particular geographical context) if their teachers lack the competence and/or confidence to make pronunciation part of their teaching. The need for an overhaul of pronunciation teaching in NNS teacher education is thus evident not only from a sociolinguistic, but also from a professional and a psychological point of view.

Research questions

How can pronunciation teaching in NNS teacher education be up-dated in order to

- a) equip NNS teachers with the linguistic and pedagogic knowledge and skills necessary to teach English pronunciation effectively and to make informed decisions in their teaching?
- b) help them develop a positive professional identity as English pronunciation teachers?

UPDATING PRONUNCIATION TEACHING FOR NNS TEACHERS OF ENGLISH

In order for NNS teachers of English to develop professional competence in the area of English pronunciation teaching as suggested in point a) and to maintain a positive self-image as pronunciation teachers (as suggested in point b), pronunciation teaching in NNS teacher education must include the following three components (see also Figure 1).

- 1) Pronunciation training
- 2) Education
- 3) Critical reflection

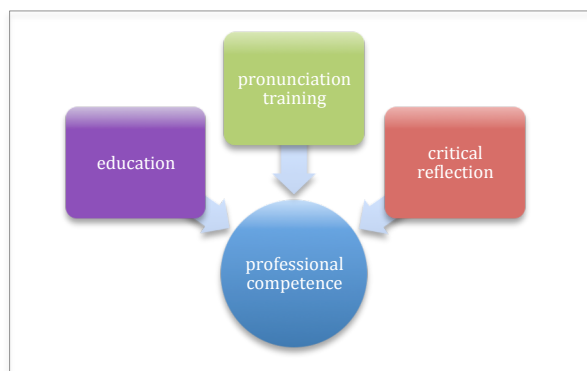


Figure 1. Suggested structure for pronunciation teaching in NNS teacher education

Pronunciation Training

In order to constitute a good phonetic role model for their future students, teachers should speak with an accent that is intelligible to a large number of interlocutors, which makes pronunciation training an indispensable component of NNS teacher education.¹¹ As the LFC constitutes a good basis for international intelligibility while still giving NNS teachers the opportunity to express their L1 identity via their accent, I suggest that the acquisition of the LFC components (including phonological accommodation skills) be made a necessary minimum requirement for *all* NNS teachers of English. This is of course not to say that student teachers who wish to go beyond the LFC and perhaps even aim for a native-like accent in English should not be allowed to do so (see also Jenkins, 2000, p. 161, 2002, p. 101). Pronunciation teaching will always have to take into account specific learner preferences and needs, be they individual or determined by local circumstances. Fortunately, nothing in the LFC is ‘unnecessary’ for learners whose goal is to acquire a NS accent (Walker, 2008, p. 9), which makes it an appropriate preliminary goal even for those who might later decide to add other pronunciation features to their accent repertoire (or who are entirely sure about wanting to acquire a NS accent right from the beginning). As Walker (2010) remarks, the LFC constitutes “an excellent foundation for learners wherever they are, and whatever their long-term pronunciation goals” (p. 46).

Education

As stated above, if English teachers are to make informed choices as to the goals and pronunciation models of their own teaching, they must be educated about the socio-psychological and sociolinguistic aspects of English pronunciation teaching (Jenkins, 1998b, p. 125). In the following, I present a number of issues that future teachers of English should be aware of and which should thus be addressed in both NS and NNS teacher education (for a more extensive list and discussion, see Thir, 2014; note that some of these points are also mentioned in Jenkins, 1998a).

¹¹ It should be noted that if NS teachers wish to act as a pronunciation model for their students, they, too, will need to speak with an accent that can be considered intelligible in a large number of contexts, and might thus have to adjust their pronunciation in some respects for the purpose of pronunciation teaching.

Phonological variation as the rule rather than the exception. Future teachers should be aware that both L1 and L2 variation in pronunciation are entirely natural phenomena, which cannot and should not be eliminated. Instead of pursuing the unrealistic (and unethical) goal of getting each and every learner to give up his/her mother tongue accent, teachers should be encouraged to take a more positive view on phonological L2 variation. That is, student teachers should understand that speaking with a foreign accent is acceptable as long as a satisfactory degree of intelligibility is maintained.

Pronunciation & identity. Future teachers must be aware that our pronunciation is “an expression of who we are or aspire to be, of how we want to be seen by others, of the social communities with which we identify or seek membership, and of whom we admire or ostracise” (Setter & Jenkins, 2005, p. 5). Given this close relationship between pronunciation and personal and social identity, individual learner preferences in L2 pronunciation learning (such as the wish to retain some features of one’s L1 accent in English) should be respected. Student teachers should understand that learners have a right to express their L1 identity via their pronunciation, as do NNS teachers.

The reality of international communication. Student teachers should be introduced to the concept of ELF and be made aware of its relevance to speakers of English in a globalized world. They must also be educated about the implications of ELF for English pronunciation teaching, i.e., the LFC and the importance of helping learners develop phonological accommodation skills. In addition, teachers should be reminded that NS norms have no direct relevance for ELF communication, and that a ‘native-like’ accent does not guarantee communicative success in ELF (as some features of NS accents might even be counterproductive to intelligibility in ELF – see Jenkins, 2000, p. 146-149; Walker, 2010, p. 41-43).

Models in English pronunciation teaching. Many teachers seem to mistake accents such as RP or GA for invariable norms that learners are to imitate regardless of the specific context of language use. Dalton and Seidlhofer (1994) contrast this notion of ‘norm’ with the one of ‘model’, i.e. a point of reference used to guide the learning process, which is “connected to language in use and therefore variable” (p. 2.7). In order to be able to use accents such as RP or GA as reference models rather than as norms, student teachers should be introduced to the distinction between model and norm according to Dalton and Seidlhofer (1994). They should be reminded that “a reference model is not ‘the truth’ or ‘the right way’ but a reference point around which many flavourings are possible” (Cauldwell, 2014).

Accentedness vs. intelligibility. Intelligibility is a complex and a relative matter, depending upon a number of different factors such as the nature of the interlocutors involved, familiarity with the accent in question, or prejudices and attitudes on the part of the listener (Rajadurai, 2007, p. 95). Future teachers should be aware of the complex nature of intelligibility, that unintelligibility does not equal accentedness (Rajadurai, 2007, p. 92), and that intelligibility is an interactive process that lies within the responsibility of both the listener and the speaker (Rajadurai, 2007, p. 90-91; Smith & Nelson, 1985, p. 333).

The role of the teacher’s accent in pronunciation teaching. As stated earlier, many NNS teachers seem to feel insecure about pronunciation teaching due to their own L1 accent in English. That is, they fail to see – and thus must be made aware – that

- 1) in a multimedia age where audio-materials of various kinds are readily available via e.g. the World Wide Web, pronunciation teaching is not solely dependent on the teacher's accent alone.
- 2) it is simply not true that only teachers with a 'native-like' accent can be good pronunciation models for their learners. In fact, Medgyes (1994) argues that NNS teachers can constitute excellent 'learner models', because they are a shining example of successful L2 learners of English. A similar point is also made by Murphy (2014, p. 259) and Jenkins (2000, p. 226), who argue that fluent NNS speakers constitute the optimum pronunciation models for international communication both in a sociolinguistic and socio-psychological sense (i.e., in terms of attainability).

A further point to which the attention of both NS and NNS teachers should be drawn is that a teacher's accent is not the only model to which learners should be exposed in the course of pronunciation teaching. Exposing learners to different teaching models is important in order to cater for the individual preferences of different types of learners, some of whom might prefer NS models for pronunciation learning while others might more strongly identify with NNS models (Murphy, 2014, p. 266). Another reason why exposing learners to different types of accents (both native and non-native) is important is the need to expand their receptive accent repertoires in order to increase their tolerance of phonological variation, and, thereby, their receptive phonological accommodation skills (Jenkins, 1998b, p. 125, 2000, p. 184, 2002, p. 100, 2005, p. 150).

Critical Reflection

Future teachers need to be encouraged to reflect critically on both traditional and novel approaches to pronunciation teaching. Thus, questions such as the advantages and disadvantages of different approaches as well as their appropriateness within different teaching contexts and for different types of learners should be considered in (NNS and NS) teacher education. In addition, NNS teachers should receive opportunities to reflect on their personal goals for English pronunciation learning: is it important for them to preserve some features of their L1 accent in English, or do they wish to become as native-like as possible? To what extent do professional or sociolinguistic considerations or dominant beliefs about the status of NNS accents in pronunciation teaching affect their personal goals? What accent would they feel most 'comfortable' with, regardless of these considerations and beliefs?

CONCLUSION

In this paper, I presented a number of suggestions how pronunciation teaching in NNS teacher education could be updated in order to equip NNS teachers with the skills needed to make informed decisions in English pronunciation teaching, especially as regards the preparation of learners for ELF interactions. This could be achieved by providing future teachers with an adequate education about the most important socio-linguistic and socio-psychosocial issues in the acquisition of English pronunciation as well as the implications of ELF for English pronunciation teaching, and by encouraging them to reflect critically on traditional and novel approaches to pronunciation teaching. A further aim of this paper was to suggest how teacher education could help NNS teachers to perceive of themselves as proficient pronunciation

teachers rather than as ‘failed native speakers’ who are incapable of teaching English pronunciation effectively. Again, education and critical reflection seem to be the key here. As regards NNS teachers’ own pronunciation skills, it was suggested that NNS teachers should, as a minimum requirement, receive pronunciation training to the extent that they attain intelligibility in international communication. However, individual preferences of NNS teachers should be respected, as all teachers should speak with an accent that they feel comfortable with. That is, as long as the criterion of intelligibility is fulfilled, it should be up to the respective teacher to decide what kind of pronunciation suits them best.

ACKNOWLEDGMENTS

I would like to thank my supervisor, Prof. Barbara Seidlhofer, for her invaluable support and suggestions for my Master’s thesis (on which this paper is based), and for encouraging me to submit an application for PSLT 2014.

ABOUT THE AUTHOR

Veronika Thir holds a teaching degree in English and French from the University of Vienna and has worked as a student tutor for practical English phonetics (focus on British English) at the Vienna English Department for several years. She combined her experience of teaching English pronunciation and her interest in English as a lingua franca for her MA thesis, in which she offers suggestions for an overhaul of English pronunciation teaching at the Vienna English Department, a major institution for English language teacher education in Austria. She is currently employed as a university assistant at the Vienna English Department as of March 2015. Her research interests include pronunciation learning and teaching, sociolinguistic variation, English as a lingua franca, and international and cross-cultural communication.

veronika.thir@univie.ac.at

REFERENCES

- Cauldwell, R. (2014, June 9). What does your accent say about you? [Web log post] British Council blog. Retrieved from <http://blog.britishcouncil.org/2014/06/09/what-does-your-accent-say-about-you/>
- Dalton, C. & Seidlhofer, B. (1994). Is pronunciation teaching desirable? Is it feasible? In S. Sebba & T. Sebba. (eds.). *Proceedings of the 4th International NELLE Conference* (2.6-2.7). Hamburg: NELLE.
- Deterding, D. (2013). *Misunderstandings in English as a lingua franca: an analysis of ELF interactions in South-East Asia*. Berlin: de Gruyter Mouton.
- Deterding, D. & Kirkpatrick, A. (2006). Intelligibility and an emerging ASEAN English lingua franca. *World Englishes*, 25(3), 391-410.
- Jenkins, J. (1998a). Rethinking phonology in teacher education. *Vienna English Working Papers*, 7(1), 40-46.

- Jenkins, J. (1998b). Which pronunciation norms and models for English as an international language? *ELT Journal*, 52(2), 119-126.
- Jenkins, J. (2000). *The phonology of English as an international language*. Oxford: Oxford University Press.
- Jenkins, J. (2002). A sociolinguistically based, empirically researched pronunciation syllabus for English as an international language. *Applied Linguistics*, 23(1), 83-103.
- Jenkins, J. (2005). Teaching pronunciation for English as a lingua franca: a sociopolitical perspective. In C. Gnutzmann & F. Intemann. (eds.). *The globalisation of English and the English language classroom* (pp. 145–158). Tübingen: Gunter Narr.
- Jenkins, J. (2006). Global intelligibility and local diversity: possibility or paradox? In R. Rubdy & M. Saraceni. (eds.). *English in the world: global rules, global roles* (pp. 32–39). London: Continuum.
- Jenkins, J. (2009). *World Englishes: a resource book for students* (2nd ed.). London: Routledge.
- Kachru, B.B. (1985). Standards, codification and sociolinguistic realism: the English language in the outer circle. In R. Quirk & H.G. Widdowson. (eds.). *English in the world: Teaching and learning the language and literatures* (pp. 11–30). Cambridge: Cambridge University Press.
- Kamhi-Stein, L., Aagard, A., Ching, A., Paik, M.S. & Sasser, L. (2004). Teaching in kindergarten through grade 12 programs: Perceptions of native and nonnative English-speaking practitioners. In L. Kamhi-Stein. (ed.). (2004). *Learning and teaching from experience: perspectives on nonnative English-speaking professionals* (pp. 81–99). Ann Arbor, MI: University of Michigan Press.
- Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, 39(3), 369-377
- Medgyes, P. (1994). *The non-native teacher*. London: Macmillan.
- Miller, J. (2009). Teacher identity. In A. Burns & J.C. Richards. (eds.) *The Cambridge guide to second language teacher education* (pp. 172–181). New York: Cambridge University Press.
- Moussu, L. & Llorca, E. (2008). Non-native English-speaking English language teachers: history and research. *Language Teaching*, 41(3), 315-348.
- Murphy, J.M. (2014). Intelligible, comprehensible, non-native models in ESL/EFL pronunciation teaching. *System*, 42, 258-269.
- Osimk, R. (2009). Decoding sounds: an experimental approach to intelligibility in ELF. *Vienna English Working Papers*, 18(1), 64-89.
- Rajadurai, J. (2006). Pronunciation issues in non-native contexts: a Malaysian case study. *Malaysian Journal of ELT Research*, 2, 42-59.
- Rajadurai, J. (2007). Intelligibility studies: a consideration of empirical and ideological issues. *World Englishes*, 26(1), 87-98.
- Seidlhofer, B. (2001). Closing a conceptual gap: the case for a description of English as a lingua franca. *International Journal of Applied Linguistics*, 11(2), 133-158.

- Seidlhofer, B. (2011). *Understanding English as a lingua franca*. Oxford: Oxford University Press.
- Setter, J. & Jenkins, J. (2005). State-of-the-art review article: pronunciation. *Language Teaching*, 38, 1-17.
- Smith, L.E. & Nelson, C.L. (1985). International intelligibility of English: directions and resources. *World Englishes*, 4(3), 333-342.
- Tang, C. (1997). On the power and status of nonnative ESL teachers. *TESOL Quarterly*, 31(3), 577-580.
- Thir, V. (2014). PPOCS 1 revisited: exploring the implications of ELF for English pronunciation teaching (MA thesis, University of Vienna). Available at <http://othes.univie.ac.at/34801/>.
- Walker, R. (2008). An achievable target for specific settings. *IATEFL Voices*, 200, 8-9.
- Walker, R. (2010). *Teaching the pronunciation of English as a lingua franca*. Oxford: Oxford University Press.

Watts, P., & Huensch, A. (2015). Assessing assessment: A principled revision of an in-house pronunciation diagnostic test. In J. Levis, R. Mohammed, M. Qian & Z. Zhou (Eds). *Proceedings of the 6th Pronunciation in Second Language Learning and Teaching Conference* (ISSN 2380-9566), Santa Barbara, CA (pp. 253-261). Ames, IA: Iowa State University.

ASSESSING ASSESSMENT: A PRINCIPLED REVISION OF AN IN-HOUSE PRONUNCIATION DIAGNOSTIC TEST

Patricia Watts, University of Illinois at Urbana-Champaign

Amanda Huensch, University of South Florida

In the last decade, interest in L2 pronunciation research and pedagogy has steadily gained momentum; yet, less attention has been paid to the area of assessing pronunciation either separately or as part of the larger construct of speaking ability. Isaacs' (2014) chapter on assessing pronunciation bemoans this fact while also noting how assessment should and could reflect recent advances in theory and research, such as the paradigm shift from accentedness to intelligibility (Levis, 2005) and findings related to intelligibility (e.g., Munro & Derwing, 2006).

In this paper, the researchers evaluate and revise an existing pronunciation diagnostic test based on a review of pronunciation assessment literature. Some modifications of the current test included the addition of a section testing aural perception, changing the free speech section from a self-introduction to an interview, and revising the targeted segmental features based on principled selection criteria. While the focus of this paper is on a test of English pronunciation within the context of international teaching assistant training, the authors believe the insights gained will be valuable and relevant for test development in other contexts as well as for other languages.

INTRODUCTION

A number of markers indicate sustained interest in L2 pronunciation and research over the past decade. *TESOL Quarterly* devoted an issue to the topic in 2005, the Pronunciation in Second Language Learning and Teaching (PSLLT) conference—established in 2009—continues to grow, and most recently, the inaugural issue of the *Journal of Second Language Pronunciation* is slated to debut in 2015. Despite these landmarks, the sentiment that pronunciation is overlooked still prevails—overlooked in terms of curricular focus in language programs (Derwing & Munro, 2005), teacher education and training (Breitkreutz, Derwing & Rossiter, 2001), and testing (Isaacs, 2014). With regard to testing, Isaacs (2014) notes that while the subject of L2 pronunciation teaching “conjures up images of neglect,” in comparison, L2 pronunciation testing does not even have a body of literature to document its current state (p. 142). Yet, the testing and evaluation of learners' pronunciation is an integral part of the teaching and curriculum development process (Celce-Murcia, Brinton, Goodwin & Griner, 2010) and—as Isaacs states—should reflect recent advances in theory and research, such as the paradigm shift from accentedness to intelligibility (Levis, 2005) and findings from the intelligibility literature (e.g., Munro & Derwing, 2006).

The aim of this study is to review literature related to L2 pronunciation testing and the aforementioned pronunciation research areas for the purpose of evaluating and revising an existing pronunciation test used in a university-level stand-alone pronunciation course. Pronunciation tests serve one of three purposes: 1) they can be used as a diagnostic to identify the specific features with which a student needs help; 2) they can measure achievement by

determining if a pronunciation feature has been learned; or 3) they can be part of the larger construct measuring overall oral proficiency (Harding, 2012). The test in this study is a diagnostic test (purpose 1) and is used in a course for international graduate students striving for a more intelligible pronunciation in order to meet the campus oral proficiency requirements to become teaching assistants. As such, the students are motivated learners and want to make noticeable improvements over the course of the semester. The diagnostic test administered during the first week of the course provides valuable information to a) help create individualized student learning plans by providing each student with a list of pronunciation features to improve and b) help with overall curriculum planning in terms of common problem areas shaped by the class members. The relatively high stakes nature of the teaching context and the centrality of the diagnostic to the course itself motivated us to ensure the grounding of the test on current literature and research findings. Based on a review of the literature, the researchers developed the following questions related to test components, pronunciation features, and rating scales to guide the revisions of the diagnostic test:

- 1) What are the values and limitations of read-aloud tasks and free speech tasks?
- 2) Which suprasegmental and segmental features should be selected for testing?
- 3) How can a test address the different problem areas of speakers from multiple language backgrounds equally well?
- 4) Is testing aural perception an important component of diagnostic testing?
- 5) How can rating scales reflect the paradigm shift from nativeness to intelligibility?

In the following sections, the authors summarize relevant findings from the literature and discuss the modifications made to the pronunciation diagnostic test that resulted.

RESULTS

What are the values and limitations of read-aloud tasks and free speech tasks?

One of the hallmark features of traditional pronunciation diagnostic tests is the inclusion of a read-aloud section to test learners on a variety of potential problem areas, including segmental and suprasegmental features. Read-aloud items are typically a series of sentences or a passage filled with potential problem areas that the student reads aloud. Advantages of including read-aloud passages are many. Firstly, the use of a reading passage limits the influence of other variables such as fluency, grammatical accuracy, etc. in that students are reading rather than producing their own language (Madsen, 1983; van Weeren & Theunissen, 1987). In addition, because all students read the same passage, they provide comparable speech samples for assessment purposes (van Weeren & Theunissen, 1987). Finally, the read-aloud passages can be designed to capture and/or highlight pronunciation features that might not occur as frequently in free speech but that are known to cause difficulties for learners, such as certain consonant cluster configurations (e.g., #sC) or intonation patterns (e.g., choice questions) (Celce-Murcia et al., 2010). On the other hand, there are limitations to using read-aloud passages. One of the major limitations is that reading ability becomes an ‘intervening’ variable in that reading may not reflect a test taker’s pronunciation in spontaneous speech (Koren, 1995). In fact, the oral reading ability of literate native speakers is not universal (Celce-Murcia et al., 2010). Finally, Celce-

Murcia et al. (2010) recommends that dialogues or other conversational texts be used rather than passages for testing suprasegmental features such as intonation and prominence.

In order to overcome some of the limitations of read-aloud passages, free speech samples may also be obtained in pronunciation diagnostic tests. These sections might include a prompt or series of prompts eliciting extemporaneous speech, such as picture story narration, role plays, or interview questions. While free speech prompts contain some of the limitations that are not an issue for read-aloud tasks (students may avoid difficult targets and rating can be influenced by other variables, such as grammatical accuracy), they allow pronunciation performance on tasks that are more reflective of real world communication. Depending on the type of free speech task included, an additional advantage can be invoking interaction, such as through the inclusion of collaborative tasks and paired speaking tasks (Isaacs, 2014; Koren, 1995). One of the main critiques of using free speech tasks is that ratings can be influenced by difficulties in other areas (fluency and grammar); however, it is the case that free speech *is* ratable and can receive a score (Buck, 1989). To avoid scores on free speech tasks being influenced by other variables such as grammatical errors and hesitation phenomena, raters can be trained to ignore fluency and accuracy errors not relevant to the elements being tested (Koren, 1995).

Before and After

Overall, the findings from our literature review and our practical experience led to the conclusion that the test should include *both* read-aloud and free speech tasks because these two types “complement each other” and confirm areas of difficulty (Celce-Murcia et al., 2010; Isaacs, 2014). Our original diagnostic test included a reading passage adapted from Celce-Murcia, Brinton, and Goodwin (1996) and a free speech section that contained a self-introduction prompt. Based on our findings, we made two major modifications (see Appendix A). Firstly, in the read-aloud section, we expanded our focus on suprasegmental features. Our original passage contained 33 suprasegmental targets (in comparison to 95 segmental targets) focused on intonation, contractions, linking, h-elision, and vowel reductions. It did not include any targets for prominence or lexical stress. In our revised test, we increased the number of suprasegmental targets to 80 (25 of which focused on prominence and lexical stress) by modifying the original passage and including a dialog. The inclusion of a dialog allowed us to more easily add targets focused on prominence and intonation (Celce-Murcia et al., 2010). Secondly, we changed our self-introduction to an interview so that it was more interactional and better reflected what was expected of our students in their teaching contexts.

Which suprasegmental and segmental features should be selected for testing? And, how can a test address the different problem areas of speakers from multiple language backgrounds equally well?

One source of information that should inform the decision about which pronunciation features to include is the literature on intelligibility and comprehensibility. *Intelligibility* is the “extent to which a native speaker understands the intended message” (Derwing & Munro, 1997, p. 2), and *comprehensibility* means “how difficult or easy an utterance is to understand” (Derwing & Munro, 1997, p. 2). Interest in the types of pronunciation errors that impact intelligibility the

most has been strong. Studies have noted the impact of both non-standard suprasegmental and segmental features to intelligibility and comprehensibility.

In terms of suprasegmentals, Munro and Derwing (1995) linked prosodic aspects of speech to raters' perceptions of comprehension. Hahn (2004) found that prosody—especially prominence at the sentence level—affected both overall comprehensibility and native speakers' reactions to the accent. Incorrect lexical stress also contributes to decreased intelligibility as has been noted by a number of scholars (Benrabah, 1997; Zielinski, 2008). In terms of testing, Koren (1995) noted the significance of stress and intonation at the phrase and word level. While Jenkins (2000) has presented evidence against including features of blended and connected speech in the lingua franca core and even in some native speaker settings, she does note three specific situations in which the addition of a full range of suprasegmentals is warranted: contexts in which 1) learners who will interact primarily with native speakers; 2) learners live in an English speaking country for extended periods; and 3) learners who want to sound native-like for professional or personal reasons (p. 136). Our specific teaching context does indeed fall within the constraints offered by Jenkins as our learners live in the US, interact extensively with native speaking students, and some—though not all—desire to sound native-like.

As for segmentals, data that would establish a rank order of segmentals according to their impact on intelligibility has not emerged, but Zielinski (2008) found that segmental errors (both consonants and vowels) in strongly stressed syllables impacted intelligibility the most. Additional information to help prioritize segmentals for the purpose of testing (and teaching) exists, however. Lado (1961)—in his seminal book on language testing which is still referenced today—advised testers to beware of targeting features that stand out as perceptually different but do not greatly influence understanding. One example of this might be a focus on /ð/ in function words, such as *this* or *that*. The functional load principle also yields valuable information to prioritize segmentals. Following from Brown (1991), functional load is a “gauge of the frequency with which two phonemes contrast in all possible environments” (p. 212). The more often phonemes contrast—for example the /l/ in *lap* and the /n/ in *nap*—the higher its functional load. The higher the functional load, the more important the phoneme is for inclusion in instruction (Brown, 1991; Munro & Derwing, 2006) and for testing (van Weeren & Theunissen, 1987). Lastly, given that most tests serve speakers from mixed language backgrounds, consideration of population-specific difficulties for relevant L1s is another appropriate means for selecting segmentals (van Weeren & Theunissen, 1987). Books—such as Avery and Ehrlich (1992) and Swan and Smith (2002)—enumerate common phonological errors by language background and can provide relevant information in this regard.

Before and After

The diagnostic changed considerably in terms of the pronunciation features tested. At the suprasegmental level, we added items to the read-aloud portion that tested lexical stress and prominence as well as expanded coverage of features related to blended and connected speech, such as vowel reduction in common function words, linking, and elision. At the segmental level, we were able to cull the number the segmentals significantly—from thirty to eighteen—by following the functional load principle and primarily focusing on segmentals that are problematic for the two most prevalent L1s in the course: Chinese and Korean. A partial list of the selected

segments follows: consonant clusters (e.g., /fl/), voicing of stop sounds, /p, f/, /r, l, n/, /w, v/, /ε, æ/, /Λ, α, ɔ/, and most tense-lax vowel contrasts.

Is testing aural perception an important component of diagnostic testing?

In addition to including components that test the oral production of learners, it is also necessary to consider their aural perception. Some examples of test items that evaluate the test taker's ability to perceive sounds or differences between sounds would be those that test sound to graphic symbol perception, dictation, or minimal pairs. Learners might be asked to accurately distinguish minimal pairs or indicate the syllable in a word that received lexical stress.

Perception and production of pronunciation are widely recognized as different skills (Koren, 1995; Lado, 1961). Testing listening discrimination skills is important because they are part of the development process for oral production (Celce-Murcia et al., 2010). Testing perception has the additional benefit of allowing a teacher to discover if test takers' difficulties are related to perception, production, or spelling (Isaacs, 2014). Ultimately, we wanted to develop a pronunciation diagnostic that systematically tests perception and production at sound, word, and phrase levels (Isaacs, 2014).

While there are an overwhelming number of advantages of testing aural perception skills on a pronunciation diagnostic exam, there is also a downside to consider. Namely, diagnostic tests must be logistical and practical in addition to valid and reliable (Lado, 1961). Testing perception increases exam time and grading load. Ultimately, however, we concluded that perception tasks should be included to provide a more complete picture of learner oral development given that perception is a related, but separate component of oral production (Celce-Murcia et al., 2010; Koren, 1995; Lado, 1961).

Before and After

Our original diagnostic did not include explicit perception items because we thought doing so would greatly lengthen the time needed to test each student individually. However, given the importance of testing aural perception skills, both segmental and suprasegmental sections were added, making perception a full-formed part of the test (see Appendix B). Targets were chosen based on the selection principles from the previous section. Segmental perception included ten discrimination items and ten identification items, and suprasegmental perception included ten lexical stress items, nine prominence items, and nine reduction items.

How can rating scales reflect the paradigm shift from nativeness to intelligibility?

Isaacs (2014) has noted that the paradigm shift away from accentedness and toward intelligibility should impact rating scales and the language they contain. To that end, she recommends the following:

- Explicitly defining terms raters may interpret differently—such as *pronunciation*, *comprehensible*, *intelligible*

- Including references to specific error types to ensure the scales are informed by the linguistic factors that lead to comprehensibility, for example, including information about sentence level prominence
- Avoiding relativistic descriptors, such as “basically unintelligible”

In addition to scales, another foundation of reliable rating is training and norming (Celce-Murcia et al, 2010; Isaacs, 2014). Specific advice includes the need to guide raters on how qualities manifested in test takers’ performance align with scale levels (Isaacs, 2014), for example, the prominence (or lack thereof) in a speech sample and its impact on overall comprehensibility. In terms of assessing diagnostics, van Weeren and Theunissen (1987) advise instructors to keep a written record noting specific pronunciation errors while scoring. In our own experience, we have found it helpful to listen to a speaker’s performance on the read-aloud section first, taking notes, and then listen to see if performance on the free speech sample confirms or disconfirms the initial judgment. Raters, especially those who are less experienced, also benefit from knowing in advance that a speech sample will require multiple listenings.

Before and After

Our original test did not include a holistic rating scale, nor did students receive a global performance rating. While it is possible to rate diagnostic samples in this manner, we chose not to in order to avoid any confusion or conflict with the holistic ratings our students received on the campus-wide ITA test—a test administered outside of our teaching context and not rated by the instructors. The importance of maintaining a united front and not sending a message which students might perceive as contradictory was paramount. In place of a holistic score, students received a checklist noting performance in key areas, a prioritized list of pronunciation features they needed to improve the most, and links to relevant information and resources to guide their study.

A rater training session using speech samples from student performance was already in place. No significant changes were made to rater training other than to update the training samples to reflect the new test.

DISCUSSION

While the original version of our diagnostic test worked effectively, it was not grounded in the research literature. The new version needed a trial run to test its functionality. Instructors using the new version during the fall 2014 semester found that it provided accurate and useful information about student performance and offered no suggestions for improvement. It should, however, be noted that the instructors were new to the teaching context and had no previous experience with the old version of the test. The concern that adding a perception section would greatly lengthen the testing time was unfounded, as all of parts of the test were easily completed within the allotted fifteen minute time slot.

Done well, diagnostic testing requires a significant investment, especially in terms of rating and providing feedback. This investment will yield the greatest return if diagnosis is seen in the larger framework of learning-oriented assessment, which aims to support learning through

feedback. To that end, providing students with diagnostic performance results, priorities for improvement, and resources for independent study sets the learning process in motion. Continued feedback and evaluation can help measure progress made on a specific pronunciation feature or alert students and teachers where additional time or effort is needed for improvement.

We have the good fortune to teach a stand-alone pronunciation course in which it makes sense to thoroughly diagnose students' pronunciation strengths and weaknesses. The test we developed is specific to our context, but we feel the advice provided here is generalizable to other contexts and languages if modified accordingly. More specifically, care needs to be taken to ensure the test fits with students' overall proficiency level and is geared toward the kinds of speaking the students will be expected to do. Even in the case of focusing on pronunciation as part of overall proficiency, Madsen (1983) recommends testing a few pronunciation targets, especially if students can be retested on those features again to measure improvement.

Based on our reading, we discovered a growing body of literature to guide the pronunciation test development/revision process, a trend that we hope continues so that a greater understanding of how pronunciation relates to overall oral proficiency and comprehensibility can be understood.

ACKNOWLEDGEMENTS

We would like to thank Sue Ingels and Veronica Sardegna for their help in creating the original diagnostic as well as Marissa Barlaz, Chelsea Coronel, and Christine Wingate for their feedback on the current version.

ABOUT THE AUTHORS

Patricia Watts is the Coordinator of the International Teaching Assistant program at the University of Illinois at Urbana-Champaign. Her areas of interest include ITA training, oral fluency, materials development, and technological applications for the teaching of pronunciation. Patricia Watts, University of Illinois at Urbana-Champaign, 4080 Foreign Languages Building, 707 S. Mathews Ave., Urbana, IL 61801, (217) 333-1506, pawatts1@illinois.edu

Amanda Huensch earned her Ph.D. and MATESL from the University of Illinois at Urbana-Champaign. She is currently Assistant Professor of Applied Linguistics at the University of South Florida. Her research interests include the acquisition of second language phonology, L2 fluency development, and ESL/TESL pedagogy. Amanda Huensch, University of South Florida, CPR 419, 4202 E. Fowler Ave., Tampa, FL, 33620, (813) 974-2548, huensch@usf.edu

REFERENCES

- Avery, P., & Ehrlich, S. (1992). *Teaching American English pronunciation*. Oxford: Oxford University Press.
- Benrabah, M. (1997). Word stress: A source of unintelligibility in English. *IRAL*, 35, 157-165.
- Breitkreutz, J., Derwing, T. M., & Rossiter, M. J. (2001). Pronunciation teaching practices in Canada. *TESL Canada Journal*, 19, 51-61.

- Brown, A. (1991). Functional load and the teaching of pronunciation. *Teaching English pronunciation: A book of readings*. New York: Routledge.
- Buck, G. (1989). Written tests of pronunciation: Do they work? *ELT Journal*, 43, 50-56.
- Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (1996). *Teaching pronunciation: A reference for teachers of English to speakers of other languages*. Cambridge, UK: Cambridge University Press.
- Celce-Murcia, M., Brinton, D. M., Goodwin, J. M., & Griner, B. (2010). *Teaching pronunciation: A course book and reference guide* (2nd ed.). Cambridge, UK: Cambridge University Press.
- Derwing, T. M., & Munro, M. J. (1997). Accent, comprehensibility and intelligibility: Evidence from four L1s. *Studies in Second Language Acquisition*, 19, 1-16.
- Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly*, 39, 379-397.
- Harding, L. (2012). Pronunciation assessment. In C. Chapelle (Ed.), *Encyclopedia of Applied Linguistics* (pp. 1-6). New York: Blackwell Publishing.
- Hahn, L. (2004). Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly*, 38, 201-223.
- Isaacs, T. (2014) Assessing pronunciation. In A. J. Kunnan (Ed.), *The companion to language assessment* (pp. 140-155). West Sussex, UK: Wiley-Blackwell.
- Jenkins, J. (2000). *The phonology of English as an international language*. Oxford, UK: Oxford University Press.
- Koren, S. (1995). Foreign language pronunciation testing: A new approach. *System*, 23, 387-400.
- Lado, R. (1961). *Language Testing: The construction and use of foreign language tests*. London, UK: Longman.
- Levis, J. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, 39, 369-377.
- Madsen, H. S. (1983). *Techniques in testing*. Oxford, UK: Oxford University Press.
- Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 49, 285-310.
- Munro, M. J., & Derwing, T. M. (2006). The functional load principle in ESL pronunciation instruction: An exploratory study. *System*, 34, 520-531.
- Swan, M., & Smith, B. (2002). *Learner English: A teacher's guide to interference and other problems*. Cambridge, UK: Cambridge University Press.
- van Weeren, J., & Theunissen, T. J. J. M. (1987). Testing pronunciation: An application of generalizability theory. *Language Learning*, 37, 109-122.
- Zielinski, B. W. (2008). The listener: No longer the silent partner in reduced intelligibility. *System*, 36, 69-84.

Appendix A

Example Diagnostic Items (selected portions of read-aloud and free speech tasks)

Directions: Review the dialogue below. Rehearse by reading aloud once or twice. When you are ready, begin.

Ben: So, how was your meeting with your new advisor?
Liz: Very informative, actually
Ben: Oh, really?
Liz: There's a good probability that I'll be a TA for him next semester.
Ben: Oh, that's great. What'll you do?
...

Part 2: Free Speech Task

Example Interview Questions

1. What do you study? Why are you in that field?
2. What problems do you think you have with oral English (oral communication)?
3. What do you hope to improve this semester?

Appendix B

Example Diagnostic Items (suprasegmental perception)

Directions: Indicate the word that receives the most stress/focus in each sentence.

Example: Where are you going? Answer: going

Dialog:

A:	Where are you studying?	_____
B:	Champaign, Illinois.	_____
A:	Do you like it there?	_____
B:	Oh, definitely. Except the winter.	_____

Brinton, D. (2015). Walk __ By: Raising learner consciousness about unstressed words (Teaching Tip). In J. Levis, R. Mohammed, M. Qian & Z. Zhou (Eds). *Proceedings of the 6th Pronunciation in Second Language Learning and Teaching Conference* (ISSN 2380-9566), Santa Barbara, CA (pp. 262-266). Ames, IA: Iowa State University.

TEACHING TIP

WALK __ BY: RAISING LEARNER CONSCIOUSNESS ABOUT UNSTRESSED WORDS

Donna M. Brinton, Educational Consultant, Beverly Hills, CA
(dmbrinton@gmail.com)

BACKGROUND

In English, unlike many of the world's languages, stressed elements occur at relatively regular intervals in a phrase or thought group, as in the following examples:

I'd like to buy a laptop.

He wouldn't eat his pancakes.

She didn't close the window.

This alternating rhythmic pattern of English is familiar to most native speakers of English due to the reliance of English poetry on rhythmic patterns such as iambic (da-DUM, da-DUM, da-DUM) or trochaic (DUM-da, DUM-da, DUM-da). In terms of pedagogical practice, the strongly rhythmic nature of English lends itself to classroom activities such as clapping or tapping in time with the stressed elements.

Stress typically falls on content words (e.g., nouns, verbs, adjectives, and adverbs), as in the sentence “Did you go to the party on Saturday?” All other elements of the phrase (e.g., articles, conjunctions, prepositions, particles) tend to be unstressed. These unstressed elements, known as function words, help to maintain the stressed-unstressed rhythmic pattern of English:

˘ ˘ , ˘ ˘ , ˘ ˘ , ˘ ˘
Did you go to the party on Saturday ?

While stressed elements are relatively easy for learners to hear, unstressed elements are much more difficult, especially for those learners from language backgrounds where stress is more evenly allocated on each syllable. Lack of knowledge about the stress-timed nature of English can lead to learner difficulties in comprehension; it also reinforces the tendency of learners from syllable-timed language backgrounds to place stress on the wrong elements of the phrase when speaking, which can negatively impact the learners' comprehensibility.

Why it's important to teach

Lack of knowledge about the phrasal stress patterns of English can result in issues of both reception and production:

1. Reception: It can interfere with learner' comprehension of the incoming stream of speech.
2. Production: It can reinforce the tendency of learners from syllable-timed language backgrounds to place stress on the wrong elements of the phrase when speaking

Teaching tip: Using a cloze passage and song

This simple-to-prepare activity involves the use of a song and a cloze passage; it is designed as a follow-up to a lesson in which students have been taught the basics of stress and unstress in English. The activity is premised on the idea that having learners focus their attention on unstressed words helps to raise consciousness about the nature of unstressed words and stress timing in English. It bears mentioning that while the typical pedagogical application for cloze passages involves deleting every n^{th} word of a song (e.g., every 5th word, or every 7th word) to supply focused listening practice, in this activity the cloze procedure has been appropriated for the teaching of pronunciation—i.e., with the cloze passage prepared in such a fashion as to eliminate the unstressed elements of the lyrics.

Key to the success of this activity is locating a song that is appropriate to the level of the learner population and that will be motivating for the learners. It is also important that the lyrics are easy to understand but not overarticulated (songs where the singer “speaks” the lyrics work particularly well). In the activity outlined below, the song “Walk on By” was selected. Originally, I intended to use the version recorded by Dionne Warwick. However, on closer listening I noticed that some of the words that would normally be unstressed are stressed in Warwick’s version. So instead, I selected the version of the song recorded by Seal. The procedures for the activity follow:

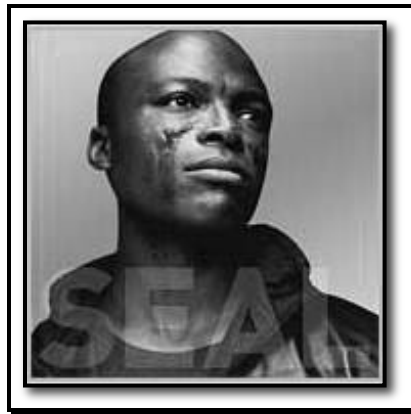
Procedures

1. Show students a photo of Seal (Appendix 1) and ask if they are familiar with the artist.
2. Elicit details about the artist and his work.
3. Distribute the handout with the cloze version of the song lyrics (Appendix 2). Explain to students that they will be listening for the missing words.
4. Allow students a few minutes to study the lyrics. (You may wish to have them predict the words that belong in the blanks.)
5. Play the song (more than once, if students want you to); have students write the words they hear in the blanks.
6. Go over the answers with the students. Ask which words were particularly difficult for them to distinguish.
7. Assuming that students enjoy the activity, the teacher can use a second song for additional listening practice. See Appendix 3 for other suggested songs.
8. As a possible extension activity, have students work in small groups to locate a song that they like. Have them prepare a cloze passage and present the song to the class following the above-outlined steps.

Useful references

- Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (with Griner, B.) (2010). *Teaching pronunciation: A reference for teachers of English to speakers of other languages* (2nd ed.). New York, NY: Cambridge University Press.
- Rogerson-Revell, P. (2011). *English phonology and pronunciation teaching*. New York, NY: Continuum.
- Teschner, R. V., & Whitley, M. S. (2004). *Pronouncing English: A stress-based approach with CD-ROM*. Washington, DC: Georgetown University Press.

Appendix 1 – Photo of the artist Seal and Cloze passage of lyrics to the song “Walk on By” as recorded by Seal



If ____ see me walking down ____ street
____ I start to cry, each time ____ meet
Walk on by, walk on by
Make believe ____ you don't see the tears
Just let me grieve ____ private
____ each time ____ I see you I break down ____ cry
Walk on by, walk on by, ____ walk on by
'Coz I ____ can't get over losing you
And ____ I seem broken ____ blue
Walk on by (just walk on by), walk on by (just walk on by)
Foolish pride is all ____ I have left

So let me hide the tears ____ all the sadness
____ you gave me when ____ said goodbye
When ____ walked on by (don't stop)
When ____ walked on by (don't stop)
____ walk on by, ____ walk on by

Walk on by (just walk on by), walk on by (just walk on by)
Foolish pride is all ____ I have left
So let me hide these tears ____ all the sadness
____ you gave me when ____ said goodbye
Just walk on by (don't stop)
Just walk on by (don't stop)
Now you really ____ ____ go so walk on by
Said you really ____ ____ go so walk on by
Baby leave me, never see the tears ____ cry (don't stop)
Baby leave me, never see the tears ____ cry (don't stop)
(Don't stop) just walk on by (don't stop)
Walk on by
(Don't stop)

Music by Burt Bacharach with lyrics by Hal David

Appendix 2 - Other suggested songs for raising awareness about unstressed words

“Tom’s Diner” by Suzanne Vega
“In the Ghetto” by Elvis Presley
“Hand in My Pocket” by Alanis Morissette
“D-I-V-O-R-C-E” by Tammy Wynette
“Yesterday” by the Beatles
“Father and Son” by Cat Stevens

“Rainy Days and Mondays” by the Carpenters
“Your Song” By Elton John
“Always On My Mind” by Willie Nelson
“Coat of Many Colors” by Dolly Parton
“Cat’s in the Cradle” by Harry Chapin
“You’ve Got a Friend” by Carole King
“Hello, Goodbye” by the Beatles
“Desperado” by Linda Ronstadt
“Dream a Little Dream of Me” by the Mamas and the Papas
“Don’t Know Why” by Norah Jones
“We are Never Ever Getting Back Together” by Taylor Swift
“Jerusalem Tomorrow” by Emmy Lou Harris

TEACHING TIP

PRONUNCIATION WORKOUT!

Marsha J. Chan, Sunburst Media and Mission College

(marsha@sunburstmedia.com)

Pronunciation is a physical act!

To produce the sounds of a language, a person needs to activate muscular components relevant to the target language or dialect. Those in the head, the neck, and the face, in particular, are instrumental in the articulation of speech. The lungs also play an important role in sending sound out through the oral cavity.

The point and manner of articulation of sounds varies, sometimes in almost unnoticeable but still important ways, from language to language. The amount and duration of pulmonary pressure from the lungs through the glottis and larynx also varies among languages. Especially when learners have had years – perhaps decades – of experience moving their mouths and controlling their breath in certain ways, it is beneficial to do stretching and breathing exercises to increase strength, flexibility, and plasticity so that learning new sounds and sound patterns can develop more readily. The breathing exercises also help learners relax, calm anxieties that may arise from speaking a foreign language, alleviate distractions, and center their minds on the here and now.

Pronunciation workout exercises can enhance a speaker's delivery in ways analogous to people engaged in athletic and musical performance. Ballet dancers' exercising regularly at the barre helps them build the balance, swiftness, and range of motion to dance gracefully and fluidly. Pianists' practicing scales and chords at varying tempos and degrees of pressure helps them play melodic compositions with appropriate lyricism, texture, and dynamics. Football players' sprinting through tires laid in a zig-zag pattern helps them gain the strength, agility and speed to dodge and weave through opponents on the field during a game. Similarly, second language speakers' doing exercises that involve stretching, tensing, and expanding the muscles of the vocal apparatus helps them improve their ability to modify their oral production and increase automaticity in communicating with a desired pronunciation and style of vocalization. In a pronunciation class, I tell my students to think of themselves as athletes of the mouth with me as their personal trainer!

While cognitive skills are certainly involved in conceptualizing, perceiving, forming, and categorizing sounds, this teaching tip focuses on bringing speakers in touch with the physical actions of the body to perform and modify spoken language so that they can proceed with the perception and production of specific sounds and prosodic patterns of the target language.

Pronunciation Workout Videos

If a picture speaks a thousand words, a motion picture speaks a million. What better way to present pronunciation workout exercises than by sight, sound, and kinesthesia via video? To help

readers visualize this teaching tip, I refer them to videos on [Pronunciation Doctor's](#) (the author's) Youtube Channel. This channel contains hundreds of videos curated into playlists, one of which is called [Pronunciation Workout Videos](#). Below is a sampling of workouts that learners can do to enhance performance in spoken language, such as pronunciation, oral presentation, and dramatic interpretation. Of varying lengths and complexity, some recorded in a studio and others before a live audience, these workout video clips and techniques may be used in class to jump-start a lesson, no matter what the target language is. As a teacher, you may use these videos in class as a co-presenter to lead your students, as a model on which to develop your own workout, and as a means for your students to practice on their own outside of class.

Based on comments posted by users from around the world, the workouts are also effective for individuals learning outside a teacher-fronted classroom.

- Ludmila Spivak Excellent learning experience. Exactly, what I was looking for to improve my speaking skills. THANK YOU SO MUCH dear Pronunciation Doctor, Marsha Chan, for sharing your knowledge!
- salal2007: THANK YOU!!!! what a joyful, great class! very useful, not only for non-native English speakers but anyone who needs to improve their vocal expression.
- Marina Babichova: you are hilarious! and it's a very useful and fun warm up! thanks!
- Alexis Granizo: You the best!!! thanks for share this exercises it helps to take my fears out.
- The Marwan: thanks that's was helpful now i can rap faster.

[Pronunciation is a physical act. Let's do warm-up exercises! \(17:40\)](#)

This video shows that pronunciation is a physical act. Warm up your vocal apparatus in a pronunciation workout! To speak, to pronounce, to modify your accent, you need to control your lips, teeth, tongue, throat and lungs. Consider me your personal trainer. Use a mirror to see yourself. Stretch your face, lips, and tongue.

Suggestion: Find time to do physical exercises 10 minutes every day to build and develop flexibility. No fees! No uniform! No equipment! Do it! Work out!

[Pronunciation Workout - Stretch your jaw and cheeks \(3:36\)](#)

In this exercise, you'll stretch your jaw wide open "AHHHHH!" /a/. Lower your jaw and tongue. Stretch your cheeks from side to side. Make sure both the upper lip and lower lip are moving. Keep your nose in front!

[Pronunciation Workout - Stretch your lips oo-ee \(3:36\)](#)

In this exercise, you'll stretch your lips far forward as you make the sound "oooooo" /uw/. Then you'll stretch your lips and cheeks as you make the sound "eeeeee" /iy/.

[Pronunciation Workout - Stretch your tongue \(5:31\)](#)

In this exercise, you'll stretch your tongue out as far as it can go. Feel the muscles in the root of your tongue. As you practice over time, stretch farther and farther. Move your tongue faster and faster. This will help you develop strength, flexibility and automaticity.

Pronunciation Workout - Waggle your tongue (4:12)

In this exercise, you'll waggle your tongue in and out. Feel the tip of your tongue hit the inside upper front teeth on the way out.

Pronunciation Workout - Poke your cheeks and clean your teeth with your tongue (5:08)

In this exercise, you'll use the tip of your tongue to poke the inside of your left and right cheeks. Then you'll stretch your tongue so that it wipes the outside upper teeth from the far right molar around the front teeth to the far left molar and back. Finally you'll stretch your tongue so that it wipes the outside lower teeth from the far right molar around the front teeth to the far left molar and back.

Pronunciation Workout - Vibrate your lips and blow through them (5:41)

In this exercise, you'll first loosen your upper and lower lips manually. Then you'll vibrate your vocal cords and close your lips as for the sound 'mmm' /m/. Feel the vibration of your lips. Next, press the tips of two fingers on the cheeks near the corners of your mouth so that your lips protrude. Vibrate your vocal cords, and with a deep breath, send air and vibration through your lips. Trill your lips. Make funny noises!

Pronunciation Workout (Standing) (8:58)

Pronunciation is a physical act. Warm up your vocal apparatus in a pronunciation workout! In this exercise, you'll stand up, stretch your arms, shoulders, torso, and back. You'll relax your head and neck. You'll stretch your mouth and jaw, expanding the oral cavity. You'll breathe deeply and send relaxation throughout your body. You'll loosen your lips and tongue saying "blah-blah-blah." You'll stretch your tongue out, increasing the distance and speed. You'll vibrate your vocal cords and trill your lips. You'll listen and clap different rhythms.

CONCLUSION

The pronunciation workout exercises are designed to help learners become aware of the vocal apparatus and make the muscles stronger and more flexible so as to enunciate more clearly and modulate oral production with greater ease. A new exercise may be introduced each session, or a variety of exercises can be led, depending on time, learners' interest and motivation, teacher's comfort in presentation, and the accessibility of the video clips on YouTube.

Further Resources

Chan, M. (1994). *Pronunciation Warm Up*, in K. Bailey & L. Savage, (Eds.), *New ways in teaching speaking*, (pp. 202-204). Alexandria, VA: Teachers of English to Speakers of Other Languages, Inc. (TESOL).

Chan, M. (1985). Video: *Marsha Chan's pronunciation workout for foreign language learners*. APAAC Video Group: Sunnyvale, CA.

Henrichsen, L. (2015). Video voiceovers for helpful, enjoyable pronunciation practice (Teaching Tip). In J. Levis, R. Mohammed, M. Qian & Z. Zhou (Eds). *Proceedings of the 6th Pronunciation in Second Language Learning and Teaching Conference* (ISSN 2380-9566), Santa Barbara, CA (pp. 270-276). Ames, IA: Iowa State University.

TEACHING TIP

VIDEO VOICEOVERS FOR HELPFUL, ENJOYABLE PRONUNCIATION PRACTICE

[Lynn Henrichsen](#), Brigham Young University

In a technology-based, entertaining way, video voiceovers combine the advantages of rehearsed pronunciation practice and tracking. This Teaching Tip report explains the pedagogical theory and procedures associated with video voiceovers, as well as the benefits that video voiceovers have produced in the author's ESL pronunciation class. Video examples of students preparing and presenting their video voiceovers (some of which were shown in the actual conference presentation) are accessible via YouTube links listed in appendices to this report.

INTRODUCTION

Rehearsed speaking activities, such as memorized speeches, provide useful pronunciation practice that can focus English learners' attention on accuracy and fluency (Morley, 1991). Research has shown the benefits of systematic, controlled imitation and repetition for improving learners' accuracy, fluency, and complexity in L2 speaking skills. Initially, in the short term, imitation activates vocabulary and syntactic structures. In the long term, repetition leads to the proceduralization of speaking skills (Davy, 2013; de Jong and Perfetti, 2011; Lynch & Maclean, 2000; Segalowitz & Segalowitz, 1993). Nevertheless, rehearsed, repetitive activities can also be boring for students—during both the practice and presentation stages. This boredom reduces student motivation and enjoyment.

Tracking is an activity in which English language learners listen to and analyze an audio or video model of speech and then imitate and reproduce it, speaking the same words simultaneously with the model. Tracking and similar procedures¹² have been recommended (Celce-Murcia, Brinton, & Goodwin, 1996, p. 310; Celce-Murcia, Brinton, Goodwin, & Griner, 2010, p. 342; Hardison & Sonchaeng, 2005; Monk, Lindgren, & Meyers, 2004; Rosse, 1999) as useful for building English language learners' fluency and increasing their awareness of prosodic features, such as pitch range, intonation, rhythm, and pausing.

Video voiceovers for L2 pronunciation practice, as described and explained in the remainder of this report, combine the advantages of tracking with those of rehearsed, repetitive activities in a fun, entertaining way that increases learners' enjoyment and motivation in pronunciation classes.

¹² Similar to tracking are procedures called mirroring and shadowing. Mirroring “involves repeating simultaneously with a speaker (in person or on television) while imitating all the speakers' gestures, eye movements, and body posturing.” Shadowing, also called *echoing*, is like tracking except that learners repeat “slightly *after* rather than *along with* the speakers” (Celce-Murcia, Brinton, & Goodwin, 1996, p. 310).

PROCEDURE

The typical video voiceover procedure, which takes three or four one-hour class periods, involves five steps:

1. The teacher explains the assignment and shows models of video voiceovers (see Figure 1). Appendix A contains links to a number of video examples from my advanced ESL pronunciation class, recorded during the years that I have used video voiceovers.
2. In class or as homework, students go online, search for, and choose a short video clip from a movie, television program, commercial, YouTube®, etc. The clip must involve one or more characters speaking naturally, and it should show the characters' mouths moving as they speak. It may be funny or serious, but in either case the content should conform to local standards of propriety for classroom presentations.
3. Over several class sessions, as well as outside of class, students work on their video voiceovers, either individually or in small groups. They first create or download a written script and decide who will take what part(s). After students have chosen their video clip and speaking parts, they listen to the clip multiple times—tracking, and paying careful attention to both segmentals and suprasegmentals, as well as pausing and pacing. As students do this, the teacher circulates and offers assistance as needed. Appendix B has links to a number of video examples showing students involved in these activities.
4. After they learn their parts, and can speak/track along with the original at normal speed, with appropriate timing, students mute the audio but continue speaking aloud. They then work on synchronizing their speech with the video characters' actions, gestures, and mouth movements.
5. After sufficient rehearsal time—both in and out of class—students present their clips to the entire class, speaking aloud while the muted video image is projected. (See Appendix A for video examples.)



Figure 1. Students presenting a video voiceover in class (left) and projected screen image (right).

BENEFITS

While formal research has not yet been conducted on the benefits of video voiceovers for improving students' pronunciation, experience and end-of-semester course evaluations demonstrate that video voiceovers are among the students' favorite activities in my semester-long advanced ESL pronunciation course. Students like video voiceovers because they are motivating, entertaining, and beneficial. First of all, knowing that they will present their video voiceovers to a live audience increases presenters' motivation to prepare thoroughly so they can perform at their very best.

Class members watching other students' presentations find them very entertaining. That enlivens the pronunciation course. The facts (1) that the content they are presenting is not only authentic but also inherently entertaining and was selected by the students themselves and (2) that the projected image of the video provides attractive, professional quality visual support, also make student presentations interesting.

Students find that preparing and delivering their video voiceovers in this manner helps them improve their speaking accuracy and fluency in an enjoyable yet effective way. As they practice, they work on speaking rate, rhythm, pausing, and intonation, as well as segmentals. Initially, many students struggle to get all these elements right. They can't keep up with the video model, or they don't pause in the right places. Some words are difficult to pronounce, and students' intonation and rhythm don't always match the model's. However, after multiple repetitions along with the video model (tracking and imitating it), student improvement is frequently dramatic. When students can finally speak at the model rate, with correct prosody and articulation, and in synchronization with the video image, they are very pleased (and sometimes amazed) with their progress.

While I have not yet developed a specialized rubric for evaluating students' video voiceover presentations,¹³ my co-teachers and I give students personalized feedback (compliments and

¹³ ESL 302 Advanced English Pronunciation for International Students is a pass/fail class, so most assignments are

suggestions) after their performances. One semester, we added an extra, fun dimension to this evaluation; my teaching assistant organized an “Academy Awards” activity. After all the students presented their video voiceovers, class members voted on which voiceovers were best in various categories: Funniest, Best Prepared, Most Dramatic Voice, Closest Voice to Original, Best Timing, Presentation of Multiple Personalities by a Single Speaker, etc. (We created many categories so that most, if not all, students would get an award.) After the votes were tabulated, we held an awards ceremony and presented the winners with certificates and plastic “Oscar” statues from the local dollar store. To make this awards ceremony even more fun, we dressed up in formal clothing, rolled out a fake “red carpet,” and served sparkling apple juice in plastic goblets (see Figure 2). Appendix C has links to videos taken during this awards ceremony.



Figure 2. “Academy Awards” presentation for best video voiceovers in various categories.

LIMITATIONS

To do video voiceovers successfully, students need to be at least at an intermediate level of speaking proficiency. They also need to have access to the necessary technology (e.g., computers/tablets and Internet access). The classroom also needs to have the technological hardware needed to show the video clips on a large screen while students speak their voiceovers.

Finally, it should be noted that video voiceovers are just one of a variety of practice activities that should be used in a good ESL pronunciation course. Overusing any single activity (no matter

graded only as “done” or not.

how fun it is) can diminish student interest and motivation instead of increasing them.

ACKNOWLEDGEMENTS

I would like to acknowledge the great work done by Ivan Bueno, Jasmine Carignan Dulin, Chirstin Stephens, Dayna Christensen Cuenca, Kaitlin Berggren Dunford, Kaylyn Johnston, and Jennifer Howell, who over the course of several semesters worked as teaching assistants and interns in my ESL 302 Advanced English Pronunciation for International Students class and helped my ESL students with their video voiceovers. Also, special thanks go to the ESL students in this class who not only performed their video voiceovers very well but also gave me permission to record and share them.

ABOUT THE AUTHOR

Lynn E. Henrichsen is a professor of TESOL in the Linguistics and English Language Department at Brigham Young University. He teaches courses in instructional methods (including the teaching of listening, speaking and pronunciation), phonetics/phonology, and ESL pronunciation. Contact information: 4040 JFSB, BYU, Provo, UT 84602, 801.422.2938, Lynn_Henrichsen@byu.edu

REFERENCES

- Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (1996). *Teaching Pronunciation: A course book and reference guide*. (1st ed.). New York: Cambridge University Press.
- Celce-Murcia, M., Brinton, D. M., Goodwin, J. M., & Griner, B. (2010). *Teaching Pronunciation: A course book and reference guide*. (2nd ed.). New York: Cambridge University Press.
- Davy, C. K. (2013, October). A cognitive approach to understanding and improving speaking skills in proficient L2 learners: A case for repetition and imitation. Paper presented at Second Language Research Forum (SLRF). Provo, UT.
- de Jong, N., & Perfetti, C. A. (2011). Fluency training in the ESL classroom: AN experimental study of fluency development and proceduralization. *Language Learning*, 61, 533-568.
- Hardison, D. & Sonchaeng, C. (2005). Theatre voice training and technology in teaching oral skills: Integrating the components of a speech event. *System*, 33, 593-608.
- Lynch, T., & Maclean, J. (2000). Exploring the benefits of task repetition and recycling for classroom language learning. *Language Teaching Research*, 4, 221-250.
- Monk, M., Lindgren, J., & Meyers, C. (2004, March). Documenting prosodic acquisition with the mirroring technique. Paper presented at the annual meeting of Teachers of English to Speakers of Other Languages, Long Beach, CA.
- Morley, J. (1991). The pronunciation component in teaching English to speakers of other languages. *TESOL Quarterly*, 25, 481-520.

Rosse, M. (1999). Tracking—A method for teaching prosody to ESL learners. *Prospect*, 14(1), 53-61.

Segalowitz, N. S., & Segalowitz, S. J. (1993). Skilled performance, practice, and the differentiation of speed-up from automatization effects: Evidence from second language word recognition. *Applied Psycholinguistics*, 14, 369-385.

APPENDIX A

Video voiceover class presentations

These YouTube® video clips show a number of different final video voiceover presentations in my advanced ESL pronunciation course.

Two students present a segment from the movie *Tangled*. <http://youtu.be/xOMPv-n2S9s>

Several students present a segment from the movie *Toy Story*. http://youtu.be/eegT_TizcQs

Several students present a segment from the movie *Home Alone*. <http://youtu.be/MX5P-URdKVo>

A student presents a segment from a **Progressive Insurance** commercial. <http://youtu.be/qfaUNddjcr0>

A student presents a segment from the movie *The Princess Diaries*. http://youtu.be/_P3QX8ACQXk

Several students present a segment from the movie *The Longest Yard*. The portion where one character laments, “You broked-ed my nose,” is especially good. <http://youtu.be/AuLxAVmHLXo>

Several students present a segment from an episode of the television show *Friends*, in which characters taste some **English trifle** (which apparently isn't very good) prepared by another character. <http://youtu.be/XbjuQzaOBgQ>

Two students present a segment from the movie *The Lion King*. <http://youtu.be/GGae3rnKLMg>

Two students present a segment from the movie *Kung Fu Panda*. <http://youtu.be/rn2ygcZ5e74>

Three students present a segment from the movie *Mulan*. <http://youtu.be/vxqzjZbXRZk>

Four students present a segment from a BYUTV *Studio C* comedy skit, titled “Candyland” <http://youtu.be/KbIpHuRty5U>

APPENDIX B

Sample videos of students engaged in voiceover preparation and practice during class

Two students **choose the video clip** (from *The Lion King*) that they will later present to the class. http://youtu.be/KbNuzu2N_RA

Two students explain **how they came up with the script** for their video clip (from *The Lion King*). http://youtu.be/d3weEAD_hl4

Two students explain **how they found and downloaded the script** for their video clip (from *Mulan*). <http://youtu.be/SnCRrCqgW1o>

Two students **practice with their script and video clip** (from *The Lion King*).
<http://youtu.be/eYTifdpioes>

Two students **track (speak simultaneously) with their video clip** (from *Mulan*).
<http://youtu.be/FYwcuMuNjss>

Several groups in my advanced ESL pronunciation class **practice their video voiceover presentations in small groups** before presenting them to the entire class.
<http://youtu.be/tKT1NtlkvxY>

Several groups in my advanced ESL pronunciation class (another semester) **practice their video voiceover presentations in small groups** before presenting them to the entire class.
<http://youtu.be/wJ0AOY8jlnk> ng

APPENDIX C Video voiceover class awards ceremony

The first part of our end-of-semester "**Academy Awards**" ceremony in my advanced ESL pronunciation course. <http://youtu.be/arBBQofvibc>

The second part of our end-of-semester "**Academy Awards**" ceremony in my advanced ESL pronunciation course. <http://youtu.be/aPmjBAZwTks>

Lima, E. F. (2015). Feel the rhythm! Fun and effective pronunciation practice using *Audacity* and sitcom scenes (Teaching Tip). In J. Levis, R. Mohammed, M. Qian & Z. Zhou (Eds). *Proceedings of the 6th Pronunciation in Second Language Learning and Teaching Conference* (ISSN 2380-9566), Santa Barbara, CA (pp. 277-284). Ames, IA: Iowa State University.

TEACHING TIP

FEEL THE RHYTHM! FUN AND EFFECTIVE PRONUNCIATION PRACTICE USING *AUDACITY* AND SITCOM SCENES

Edna F. Lima, Ohio University

While mastering the segmental features of English is certainly important, especially when it comes to the functional load principle (Munro & Derwing, 2006), many claim that mastering the suprasegmental features (word stress, rhythm, and intonation) of English is more crucial to improving learner comprehensibility (Isaacs, 2008; Jenkins, 2000; Kang, 2010). Rhythm, created by the alternation between stressed and unstressed syllables, is one suprasegmental that has a large impact on comprehensibility, as “learners who use incorrect rhythm patterns or who do not connect words are at best frustrating to the native speaking listener” (Celce-Murcia, Brinton, Goodwin, & Griner, 2010, p. 163).

The present teaching tip is designed to help adult learners improve their perception and production of rhythm in English using *Audacity* and sitcom scenes. By creating appealing and useful activities that can be assigned as homework, teachers give learners the opportunity to be exposed to a variety of speakers, practice more, work in a comfortable environment, and work at their own pace. Also, by using a script as the basis for the exercises, teachers make the feedback process easier and more effective.

This activity is part of a four-week online pronunciation tutor and focuses on rhythm and prompts learners to “imitate” the provided model. Before being assigned this activity, learners watch two short instructional videos on how to download and use *Audacity* to record and edit audio files. They will use *Audacity* to listen to the input and attempt to “imitate” it by recording their own speech right after hearing the input. *Audacity* will allow learners to record a new track and save both tracks (the original model and the new utterance) in one single file so that they can listen to both recordings and monitor their progress in approximating their speech to that given model. Given that this activity was primarily designed for the training of international teaching assistants providing instruction at American universities, the model used was American English. However, this activity can be adapted to include any given variety of English or foreign-accented speech.

Goals of the Teaching Tip

1. To demonstrate the importance of rhythm to native listeners.
2. To practice both perception and production of the English rhythm.
3. To help students develop self-monitoring skills by comparing input (model provided) with output (their imitation of the model).

Task 1 – Downloading, installing, and learning how to use *Audacity*

As mentioned previously, before doing the main activity presented here, learners watch one video on how to download and install *Audacity*¹⁴ and one video on how to record and edit files using this software¹⁵. This is specifically relevant if teachers plan to assign this and other activities as homework. Once learners are familiar and comfortable with recording and editing files in *Audacity*, they are able to perform any activity the teacher may assign in or outside of the classroom.

Task 2 – Watching a 90-second sitcom clip: *King of Queens*: Phone number rhythm

The next part of the activity is to have learners watch the sitcom clip (<https://www.youtube.com/watch?v=RsIS824ZyNI>) that they will “imitate” later. This clip is from the sitcom *The King of Queens*; in this scene, a conflict between Doug Heffernan (Kevin James) and Arthur Spooner (Jerry Stiller), a son-in-law and father-in-law, is caused by improper phone number rhythm. Teachers may start with very short clips (e.g., one single scene) and then increase the length of clips (e.g., two to three scenes). First, learners are asked to just watch the video and enjoy it. That is, the purpose is for them to understand what is going on between the characters. Next, they are asked to watch the video again and pay close attention to the dialogue. What is going on? What is the cause of conflict between the characters? Finally, students are asked to focus on the rhythm. Which words are stressed? Which words are de-emphasized? The assumption here is that learners have received some sort of instruction on how rhythm is created in English before this activity is performed.

Task 3 – “Imitating” the model

Now that learners are very familiar with the scene, they are asked to imitate the model so that they can practice producing the rhythm of English. Before assigning the activity to learners, however, the teacher needs to prepare the audio file for recording. The main purpose of this activity is to have learners repeat after the model “with an option for comparing input with output” (Neri, Cucchiaroni, Strik, & Boves, 2002, p. 461). After recording themselves repeating after the model, learners listen to their audio files and monitor their own speech (self-monitoring). “How well did I imitate the model”? Learners then compare their output with the input. They then can re-record the excerpt as many times as they wish until they feel satisfied with their performance.

Preparing the audio file for recording

Step 1. Importing the original file into *Audacity*

Open *Audacity* and click on “File” and then “Open” (Figure 1).

¹⁴ YouTube: <https://www.youtube.com/watch?v=sJNIPWa8LnM>. This video includes instructions on how to download and install the LAME encoder, which enables *Audacity* to export files in MP3 format.

¹⁵ This is a tutorial (<https://www.youtube.com/watch?v=oXRfmbChdWU&list=UU-O0ulXD4elRqUQ5v00AAuw>) that I created for my own students. The video focuses on how to use *Audacity* to record and edit audio files. It also demonstrates how to record one track while listening to another one.

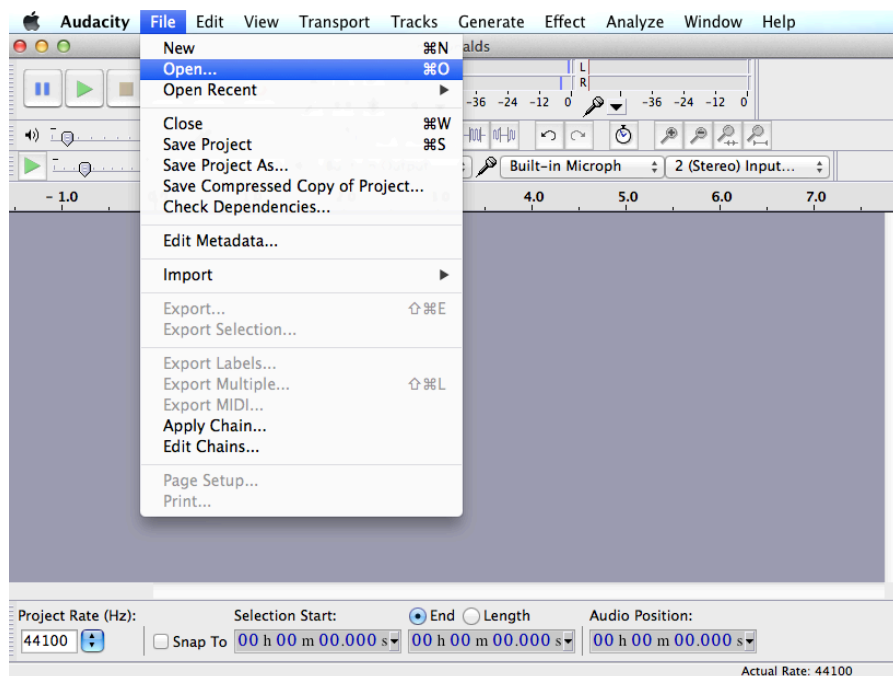


Figure 1. Importing an audio file into Audacity

Find the target file on your computer and click on “Open” (Figure 2).

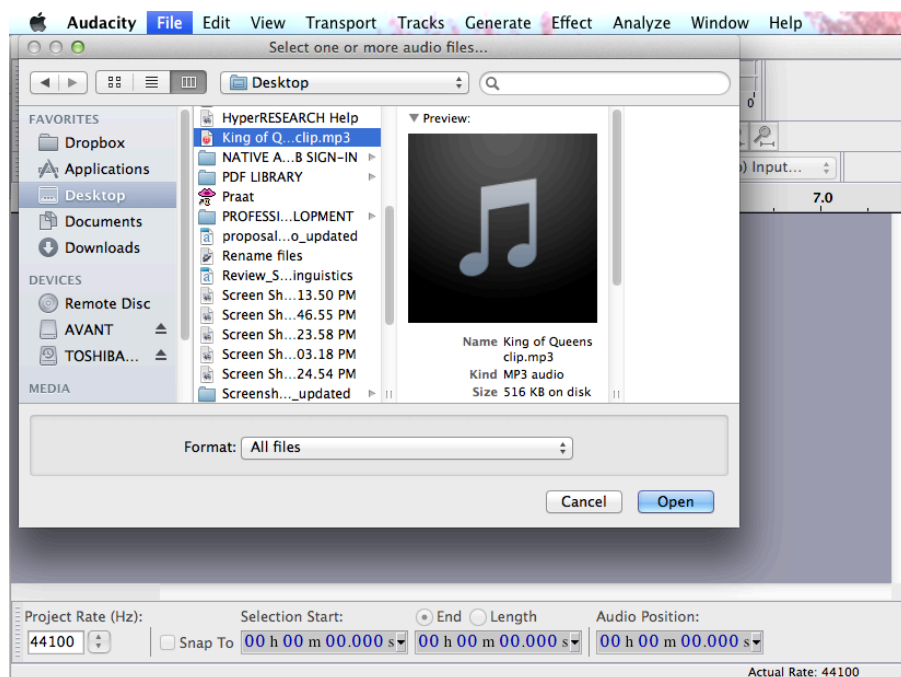


Figure 2. Selecting the audio file to import into Audacity

Step 2. Generating silence in the original file in *Audacity*

Now that the original file is open in *Audacity*, generate silence so that learners can listen to the model and repeat after each line. Place the cursor where you would like to insert silence. Click on “Generate” and then on “Silence” (Figure 3).

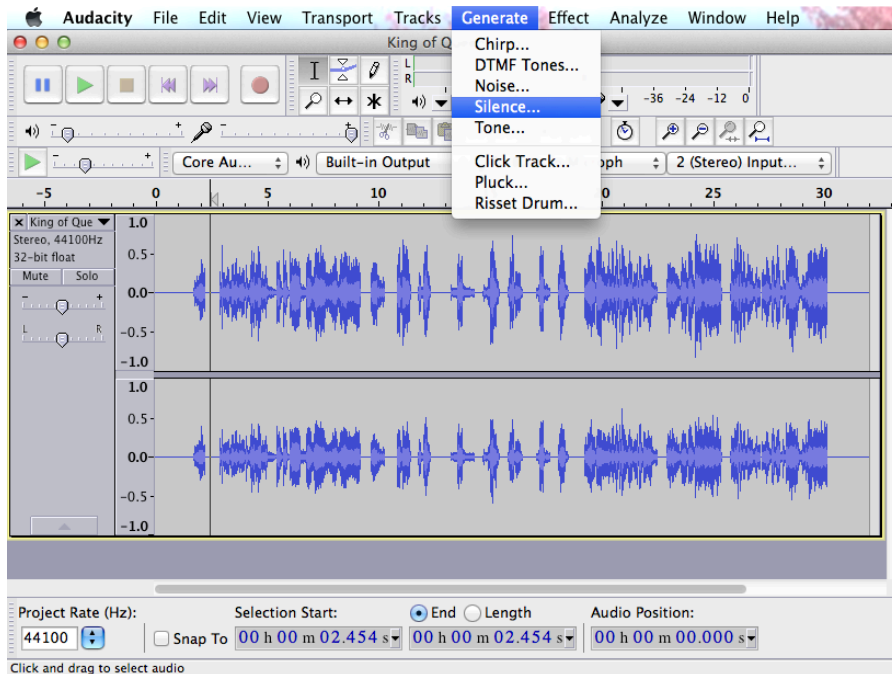


Figure 3. Generating silence between utterances in an audio file

As seen in Figure 4, enter an estimated number of seconds to allow learners enough time to repeat after the model. Click “Ok”. To avoid frustration, it is important to give learners enough time so that they are finished recording one line before the next line is played (see example below).

Example

Model: You KNOW WHAT?

Learner: You KNOW WHAT?

Model: HOW about a little PHONE NUMber RHYTHM, HUH?

Learner: HOW about a little PHOne NUMBER RHYTHM, HUH?

Model: You KNOW, a little BUM PUM PA BUM PA BUM PA...

Learner: You KNOW, a little BUM PUM PA BUM PA BUM PA...

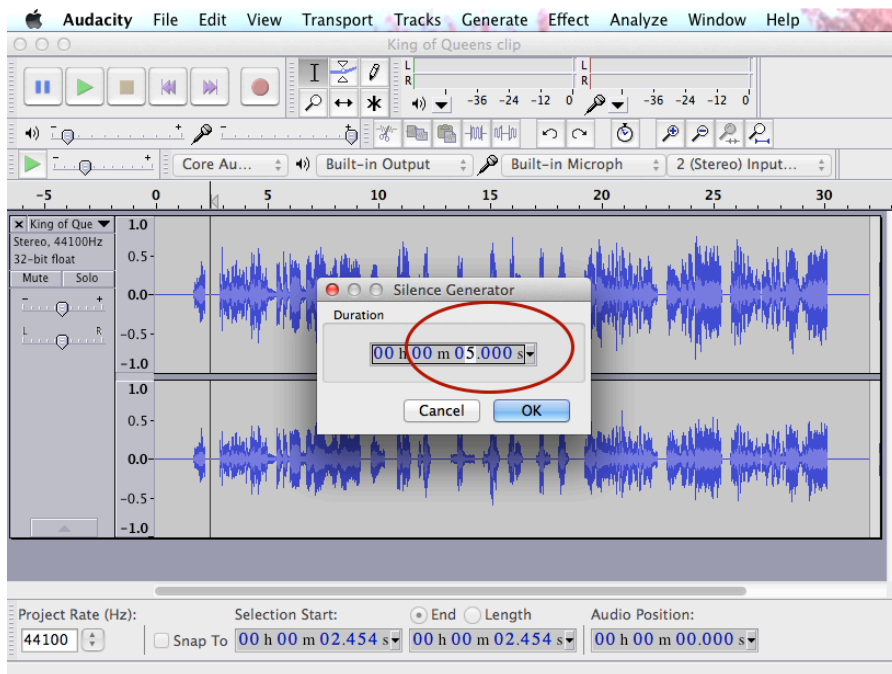


Figure 4. Generating silence depending on duration desired

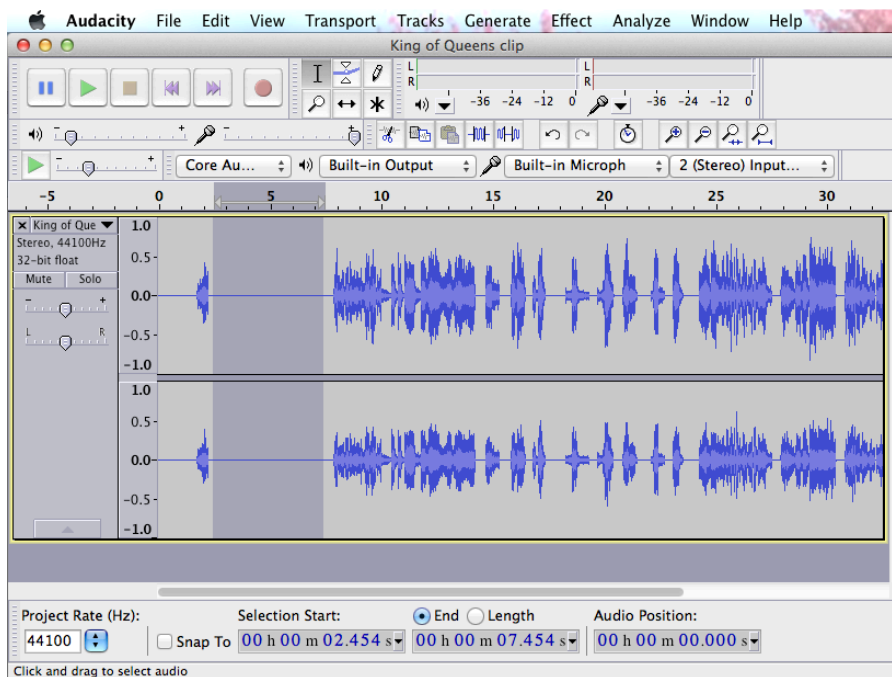


Figure 5. Five-second silence inserted after “You know what?”

Repeat step 2 until the file is ready for recording. Below are the task instructions provided to my students. The instructions were available in Moodle, and students were asked to upload their recording by a given deadline. Teachers can easily adapt these instructions in order to reflect the context in which the activity is taking place.

Rhythm Production Exercise 1

Now that you are familiar with *Audacity* (Tutorial 1: Downloading and Installing Audacity and Tutorial 2: Using Audacity to Record and Edit Audio Files) and know how to use it to record yourself, let's do a rhythm exercise. Do you remember the King of Queens: Phone Number Rhythm video clip? Well, now you will have the chance to practice some of the dialogue. Download the file McDonald's Scene Audio File and open it in *Audacity*. Listen to the recording and say each line after you hear it. Once you record the lines and are satisfied with your performance, upload the file to Rhythm Production Exercise 1 Dropbox.

Script of the lines you will hear in the recording:

Doug: You know what? How about a little phone number rhythm, huh? You know, a little bum pum pa bum pa bum pa... Got it?

Arthur: I got it.

Doug: Ok.

Arthur: One...

Doug: Yeah...

Arthur: Six...

Doug: Yep!

Arthur: ...teen

Doug (angry): I already dialed the 6. I can't go back in time and slip a 1 in!

Arthur (angry): Well, whose fault is that?

Doug: I'm making a sandwich. Why don't you order from... McDonald's?

The activity presented here works well for learners of different ages and at different levels of proficiency. Teachers can increase the challenge by selecting more complex scenes. However, because the main objective is to have learners imitate the model and not read from the script ("read" speech), it is important that the discourse is broken into smaller units like in the example provided. If the audio file contains long sentences, learners may be tempted to read from the script instead of just repeating after the model.

Other activity tips

The activity described here was developed to focus on the perception and production of English rhythm. Learners first pay attention to how rhythm is created (perception) and then record themselves repeating after the provided model (production). They then listen to their recording and compare the input with their output (self-monitoring). However, the same activity setup can be used to target any pronunciation feature. For instance, the same sitcom clip can be used in an intonation activity. Before imitating the model, learners could be prompted to listen to the dialogue between Doug and Arthur and decide if the final intonation for each utterance falls or rises. The teacher could provide the script with arrows at the end of each utterance (e.g., Doug: You know what? ↘ ↗) and have students circle the correct option.

Another activity possibility would be thought groups. I like using “The Big Bang Theory” clips to create thought group activities because of the difference in speed that each character shows. For instance, Sheldon (played by Jim Parsons) speaks really quickly as compared to some of the other characters (e.g., Amy, Sheldon’s girlfriend, played by Mayim Bialik). The point here: thought groups may vary depending on speech speed. Slower speakers tend to pause more, thus creating a larger number of thought groups. Teachers can, for example, select a scene in which three or four characters are interacting with each other, transcribe the scene, and have students watch the clip and mark the thought groups (/ for shorter pauses and // for longer pauses) in the transcript. Afterwards, students imitate the model in *Audacity*, paying closer attention to thought group division.

There are endless perception and production possibilities for practice involving *Audacity* and sitcom scenes. Teachers can create meaningful and fun activities on any given pronunciation feature for any given audience. The activities can be assigned as homework in case technology is not available in the classroom, and a discussion of the activity can be conducted in the classroom afterwards. Depending on class size and schedule, teachers can also give students individualized feedback in addition to self-monitoring. As mentioned previously, working off of a script and designing each activity to focus on one given target (e.g., rhythm) at a time, will likely make the feedback process easier and more effective.

Learners appreciate the opportunity to practice and are more likely to really engage in the activity when they are performing it in a non-inhibiting environment. They are also less afraid of making mistakes, and they are allowed to re-record themselves until they are satisfied with their performance. One of my students said, “Because the [phone number rhythm] scene was so funny, I imitated and recorded it about 100 times before I finally decided to upload my recording to Moodle”. Listening to her recording a few days later, I can say that not only did she have fun doing it, she felt and nailed the rhythm!

ABOUT THE AUTHOR

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. During her eight years as an international teaching assistant at Iowa State University, she taught a variety of academic courses to both domestic and international undergraduate and graduate students. Her research interests include CALL,

language assessment, material design, and technology applied to language learning and teaching, especially to pronunciation instruction. Contact her at Ohio University, 383 Gordy Hall, Athens, OH. 740.593.4538 • F: 740.593.2967. lima@ohio.edu

REFERENCES

- Celce-Murcia, M., Brinton, D. M., Goodwin, J. M., & Griner, B. (2010). *Teaching pronunciation: A course book and reference guide* (2nd ed.). Cambridge, NY: Cambridge University Press.
- Isaacs, T. (2008). Towards defining a valid assessment criterion of pronunciation proficiency in non-native English-speaking graduate students. *Canadian Modern Language Review*, 64(4), 555–580.
- Jenkins, J. (2000). *The phonology of English as an international language: New models, new norms, new goals*. Oxford: Oxford University Press.
- Kang, O. (2010). Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System*, 38(2), 301–315.
- Munro, M. J., & Derwing, T. M. (2006). The functional load principle in ESL pronunciation instruction: An exploratory study. *System*, 34(4), 520–531.
- Neri, A., Cucchiaroni, C., Strik, H., & Boves, L. (2002). The pedagogy-technology interface in computer assisted pronunciation training. *Computer Assisted Language Learning*, 15(5), 441–467.

Schaefer, V., & Darcy, I. (2015). A communicative approach and dialect exposure enhance pitch accent awareness by learners of Japanese (Teaching Tip). In J. Levis, R. Mohammed, M. Qian & Z. Zhou (Eds). *Proceedings of the 6th Pronunciation in Second Language Learning and Teaching Conference* (ISSN 2380-9566), Santa Barbara, CA (pp. 285-296). Ames, IA: Iowa State University.

TEACHING TIP

A COMMUNICATIVE APPROACH AND DIALECT EXPOSURE ENHANCE PITCH ACCENT AWARENESS BY LEARNERS OF JAPANESE

Vance Schaefer, University of Mississippi

Isabelle Darcy, Indiana University

INTRODUCTION

Japanese pronunciation is deceptively simple. Vowels are limited to five, and consonants generally pose little difficulty for L1 English speakers except for a few novel sounds such as /tsu/ or [ɾ]. Thus, the L1 English speaker commonly comes away with the impression that Japanese pronunciation is quite simple. Yet, there are subtle differences such as the /t/ being less aspirated in Japanese than its English counterpart or Japanese featuring /u/ rather than /u/. Additionally, words in Japanese are distinguished by their pitch-accent pattern.

Simply put, each mora (i.e., unit of Japanese rhythm smaller than a syllable) in a word bears either a low (L) or high (H) pitch. Thus, the 2-mora word [ka.ki] (note: a [.] separates each mora) can either mean 'persimmon' or 'hedge' if spoken in the LH pattern or 'oyster' if spoken with the HL one. Nevertheless, such lexical contrasts are again often glossed over in the Japanese as a Foreign Language (JFL)/ Japanese as a Second Language (JSL) classroom, setting learners up for a fall in communicative skills, especially when aggravated by issues in grammar or vocabulary.

One reason why pitch accent is neglected in the classroom may be due to the fact that the functional load of pitch accent to differentiate minimal pairs is considered to be low (14%, Shibata & Shibata, 1990). While most words can be understood through context, the wrong pitch pattern on a word will often not be understood in isolation. In fact, the following tongue twister illustrates that the 'intelligibility load' of pitch accent when interacting with segmentation is indeed quite large. Such a sentence would be very difficult to understand if not spoken with appropriate pitch accents:

[su.mo.mo.mo.mo.mo.mo.mo.mo.mo su.mo.mo.mo.mo.mo.mo.mo.mo.no.u.chi]
sumomo mo momo, momo mo momo, sumomo mo momo mo momo no uchi.

LHH H LH, LH H LH, LHH H LH H LH H HH

plum also peach, peach also peach, plum also peach also plum of inside

A plum is also (a type of) peach, a peach is also (a type of) peach, both plums and peaches are in the peach (family).

While this tongue twister may stretch the limits of the Japanese language, it accentuates the point (pun intended) of the actual functional load of pitch accent in Japanese.

Moreover, the pitch pattern of a word may even be modified when combined with other words (i.e., compounds) or in phrases [adjective + noun, noun (object) + verb], even more so than in a tonal language like Mandarin (cf., tone sandhi, where two adjacent syllables carrying tone 3 would be realized as the sequence tone 2 + tone 3). The following example illustrates these adjustments in compounds. Here, the initial L of 'airport' becomes H in the compound:

[ko.ku.sa.i] LHHH 'international'

[ku.u.ko.o] LHHH 'airport'

[ko.ku.sa.i.ku.u.ko.o] LHHH HLLL 'international airport'

(Tanimori & Sato, 2012, p. 22)

When combining words in noun/adjective phrases, the following example shows that the system is in fact relatively opaque. The word 'car' is pronounced with a different pitch-accent depending on its color:

[a.ka.i] 'red' LHH

[ku.ru.ma] 'car' LHH

[a.ka.i.ku.ru.ma] 'red car' LHHHHH

[a.o.i] 'blue' LHL

[ku.ru.ma] 'car' LHH

[a.o.i.ku.ru.ma] 'blue car' LHLLHH

(cf., Tanaka & Kubozono, 2012, pp. 75-76)

In addition to the 'functional intelligibility load,' pitch accent represents a social aspect where "Differences existing among Japanese dialects can be seen best in the accent of words" (Kindaichi, 1988, p. 56). For example, the pitch accent pattern of 'thank you' varies regionally as in [a.RI.ga.to.o] LHLLL (note: capital letters indicate the accent) in standard Japanese versus [a.ri.ga.TO.o] LLLHL in the Kansai dialect of Osaka and Kyoto (Kubozono, 2012). Indeed, some words often have the opposite pitch patterns in different varieties: in standard Japanese, LH [ha.shi] means 'bridge', and HL [ha.shi] means 'chopsticks'; in Kansai, this pattern is reversed: LH means 'chopsticks' and HL means 'bridge.'

Pitch accent is important at advanced levels for reading and listening to academic discourse and polite language. Pitch accent interacts with pausing to define thought groups, e.g.,

[a.ru.mi.ka.n. no. u.e. ni. a.ru.mi.ka.n]

LHHHH H LH H HL HLL

aluminum can of top on is located (that) mandarin orange

a mandarin orange (that) is located on top of an aluminum can

Similarly, presentations adopt a more formal, polite manner of speaking, resulting in changes in pitch accent. For example, the polite *o*-NOUN *versus* plain NOUN can differ, e.g.,

[su.shi] LH 'sushi'

[o.su.shi] LHL 'sushi' (polite)

The humble o-VERB STEM versus plain nonpast verb can also differ, e.g.,

[o.ka.ki shi.ma.su] LHH HHL 'I write'

[ka.ku] HL 'I write'

Thus, pitch accent is an important part of vocabulary and grammar, aiding learners in speaking fluency, understanding of expository discourse and cultural proficiency. In short, Japanese pitch accent – generally often overlooked in the classroom – is in fact the cornerstone of effecting greater global abilities in understanding and using the language among JFL students.

The goals of this teaching tip are twofold:

- 1) To provide the teacher with a short primer about Japanese pitch accent, and
- 2) To discuss teaching methods by offering sample lessons to help teachers embed pitch accent practice into lessons.

An ancillary goal is to advocate the teaching of dialectal differences to raise awareness by students of the importance of pitch accent

WHAT THE TEACHER NEEDS TO KNOW

Linguistic description of pitch accent in Japanese

Patterns. Despite the seeming opacity of the pitch accent system, there are rules or tendencies which can be exploited in the JFL/JSL classroom to aid learners in mastering pitch accent. The accented mora in a word in Japanese bears a high pitch followed by a drop in pitch from high to low. While knowing where the accent occurs is unpredictable, there are tendencies, and once the accent location is known, the pitch of each mora of a word can be easily assigned.

- Every mora in a word bears either a low or high pitch.
- The first mora is low unless the accent falls on that mora making it high.
- The first and second mora of a word must differ in the pitch so that when the first mora is low, the second must be high or vice-versa. The exception is a heavy first syllable, e.g., [ko.u.ko.u] 'high school' where in effect the entire first syllable [kou] can be pronounced high (Tsurutani, 2011).
- All the moras occurring between the first low mora and the accented mora are high in pitch.
- Once the pitch falls from high to low within a word, it cannot rise again.
- Moras following the accented mora are low in pitch (cf., Sugiyama, 2012, p. 6)

However, there are also unaccented words where there is no fall from high to low within the word. In Table 1, we can see the accent pattern types according to the placement or lack of an

accent. Also, some patterns appear more frequently than others (See figures in parentheses in Table 1).

Table 1

Types and Frequencies of Lexical Pitch Accent Patterns in Standard Japanese

No. of moras Pattern type	1 mora	2 moras	3 moras	4 moras
<i>heibangata</i> 'flat' (=unaccented)	<i>hi.ga</i> LH 'sun/day' (> 30%)	<i>ha.shi.ga</i> LHH 'edge' (approx. 15%)	<i>sa.ka.na.ga</i> LHHH 'fish' (approx. 50%)	<i>ko.ku.ba.n.ga</i> LHHHH 'blackboard' (< 70%) * <i>ko.u.ko.u.ga</i> LHHHH ⇨ HHHHH 'high school' when the first syllable is heavy
<i>atamadakagata</i> 'initial (head) '	<i>hi.ga</i> HL 'fire' (< 70%)	<i>ha.shi.ga</i> HLL 'chopsticks' (approx. 65%)	<i>mi.do.ri.ga</i> HLLL 'green' (< 40%)	<i>chu.u.go.ku.ga</i> HLLLL 'China' (< 10%)
<i>nakadakagata</i> 'medial'			<i>ta.ma.go.ga</i> LHLL 'egg' (< 10%)	<i>hi.ko.o.ki.ga</i> LHLLL 'airplane' (> 10%)
<i>nakadakagata</i> 'medial'				<i>mi.zu.u.mi.ga</i> LHHLL 'lake' (< 10%)
<i>odakagata</i> 'final (tail)'		<i>ha.shi.ga</i> LHL 'bridge' (< 20%)	<i>a.ta.ma.ga</i> LHHL 'head' (approx. 5%)	<i>o.to.o.to.ga</i> LHHHL 'younger brother' (approx. 5%)

* *ga* is a nominative particle (Based on Tanaka & Kubozono, 2012, pp. 58-59)

Tendencies. Tendencies also exist in terms of the frequency of pitch accent patterns as can be seen the chart above, e.g., the unaccented pattern is quite frequent in 3- and 4-mora words (Tanaka & Kubozono, 2012). Also, tendencies for pitch accent patterns exist for word types: most native Japanese words and Sino-Japanese loanwords are unaccented, but when accented, the accent is generally on the antepenultimate mora (see Kawahara, 2015 for details). This tendency is reflected in Western loanwords which also tend to be accented on the antepenultimate mora. Thus, we have words such as [BA.na.na] HLL 'banana' etc., where incidentally the accent does not necessarily coincide with the word stress of English. Exceptions include when the third mora from the end is the second part of a heavy syllable, e.g., [baa], yielding [ha.n.BA.a.ga.a] 'hamburger' and not [ha.n.ba.A.ga.a] (cf., Tanaka & Kubozono, 2012, p. 64).

Tendencies also exist for compound verbs, adjectives and nouns. For compound verbs where the first part is also a verb, the accent falls on the second syllable from the end on all non-past tense forms, e.g., [a.RU.ku] LHL + [tsu.zu.ke.ru] LHHH = [a.ru.ki.tsu.zu.KE.ru] LHHHHHL 'to continue walking' (Tanaka & Kubozono, 2012, p. 83). Patterns exist as well for other forms of compound verbs (e.g., past tense), compound adjectives, and to some extent for compound nouns.

Pitch accent also interfaces with grammar with fairly regular pitch accent placement. In the conjugation of verbs and adjectives, there are two types: accented and unaccented. Accented adjectives are more common than unaccented ones (Tanaka & Kubozono, 2012, p. 75) while accented verbs and unaccented verbs are in roughly equal proportion (p. 80). For accented verbs and adjectives, there is a shift in the place of accent, e.g.,

[ta.BE.ru] LHL nonpast 'to eat'

[TA.be.ta] HLL past 'to eat'

[ta.KA.i] LHL nonpast 'to be high/expensive'

[TA.ka.ka.t.ta] HLLLL past 'to be high/expensive'

We see regular patterns for pitch accent placement in other accented verb forms, e.g., [ta.be.SA.se.ta] (causative), [ta.be.RA.re.ta] (passive), or in both accented and unaccented verbs on the same affix, e.g., [ta.be.MA.su] (polite form) or [ta.be.YO.u] ('let's eat'). As such, with controlled exposure in the JFL/JSJL classroom, learners could learn to make educated guesses and/or lexicalize the accent placement with regularity.

Importance of pitch accent for communication

Pitch accent is important in Japanese: it constrains lexical access for native speakers (Otake & Cutler, 1999) and is processed differently in the brain according to dialect (Sato et al., 2013). Non-target-like productions create an impression of a foreign accent – more so for suprasegmentals than segmentals in Japanese (Sato, 1995). Yet, it is rarely taught explicitly in JFL classrooms while learners have difficulties acquiring native-like pitch-accent patterns implicitly (Shport, 2008). L1 English/L2 Japanese learners have difficulty separating the length correlate of English word stress when producing Japanese pitch accent (Kondo, 2007). This is problematic as vowel length is phonemic in Japanese. Anecdotally speaking, L1 English speakers oftentimes either speak with the patterns of English word stress or flat pitch patterns.

Perceptual training improves perception/production of suprasegmentals (Wang, Jongman, & Sereno, 2003) and has been shown to do so specifically for Japanese pitch accent (e.g., listening discrimination, metalinguistic knowledge, Hirano-Cook, 2011). Improved perceptual/listening abilities result in enhanced production/speaking (Rvachew, Nowak, & Cloutier, 2004). Furthermore, accurate listening comprehension and improved segmentation lead to more accurate writing (in a way similar to English for example, where better perception and segmentation lead to not leaving out function words or reduced forms, Celce-Murcia et al., 2010; cf., “*sumomo*” and “*aluminum can*” examples above for Japanese).

PRACTICAL SUGGESTIONS

Despite the significance of pitch accent in Japanese, in JFL/JSJL courses, pronunciation, let alone pitch accent, is rarely if ever taught. Instructors note that they do not have any chance to teach pronunciation nor have any confidence in teaching pronunciation (Kawasome, 2014) and that

pronunciation is often *atomashi ni naru* 'put off until last'. This paper advocates an approach based on the communicative framework prioritizing guided practice (Celce-Murcia et al., 2010), and enriched by less conventional methods such as dialect exposure.

Communicative Framework

Following the guiding principle of a simultaneous dual focus on both form and meaning (Segalowitz & Hulstijn, 2005), pronunciation should be integrated into all skill activities (e.g., grammar, reading) to boost pitch accent awareness and facilitate more target-like (lexical) acquisition. Automaticity in pronunciation necessitates repetition with attention to form (Gatbonton & Segalowitz, 1988). Yet, repetitive drills oftentimes lack context, whereas meaning-focused activities lack the repetition necessary for automaticity. The solution to this pedagogical conundrum lies in devising activities that are intrinsically repetitive in nature while simultaneously being communicative (Gatbonton & Segalowitz, 1988). That is, in order to maintain a dual focus on form and meaning, activities should ideally require that learners accurately perceive/produce the target form in order to complete the activity (Loschky & Bley-Vroman, 1993; See Sicola and Darcy (2015), for more extensive discussion).

Within a communicative framework there are five stages: descriptive/analysis (metalinguistic awareness), listening, controlled, guided and communicative practice (Celce-Murcia et al., 2010). Learners are first made aware of pitch accent, gaining metalinguistic knowledge through simple explanations or exercises, e.g., marking accent in words, and deducing patterns. In the next stage, students listen for the feature, e.g., differences between minimal pairs or pitch patterns. In the following three stages, there is a gradual shift from a heavier focus on form (i.e., pitch accent patterns) to function (i.e., communication) with practice occurring within context and with repetition where communication hinges on correct pitch accent usage. Naturally, all these steps overlap and do not always progress in a defined order.

Dialect Exposure

Exposure to non-standard dialectal pitch-accent patterns should boost metalinguistic awareness and improve perception/production of pitch accent (cf., Baker & Smith, 2010 for L2 segmentals). Simply put, exposure to dialect through hearing short, simple dialogs or extended speech where pitch patterns (along with other pronunciation features) between standard Japanese and Kansai Japanese differ could jolt students into greater awareness of pitch accent.

Examples of Activities

We present activities which can be embedded into other lessons to boost the awareness of pitch accent. Simple activities include the instructor having the students notice the pattern by merely pointing it out or having students deduce the pattern. More complicated activities require retooling typical activities or supplementing lessons with entirely new activities.

Topic: Minimal pairs

Level: Beginner

Step: Awareness, listening discrimination, controlled practice, guided practice

Also: *hiragana*, *te*-form of verbs (instructions)

Activity: Total Physical Response (TPR)

The instructor uses the *te*-form of verbs (e.g., *-te kudasai* 'please do x') to instruct students to do something, using pitch-accent minimal pairs as in 'Please draw/write rain' (on the board/paper),

‘Please touch rain’ (a picture on the board/wall/handout), etc. After the instructor models TPR, the students then practice amongst themselves with the instructor monitoring.

Topic: Minimal pairs

Level: Beginner

Step: Awareness, listening discrimination

Also: *hiragana*

Activity: Bingo

The instructor creates bingo boards with minimal pairs differing in pitch accent (e.g., *hashi*, LH and HL) and plays bingo as an entire class or in small groups. Minimal pairs varying by vowel/consonant length can be added as well. The illustration below shows an example of a bingo card using both or just one member of the minimal pairs and fillers (e.g., spider).



Topic: Minimal pairs

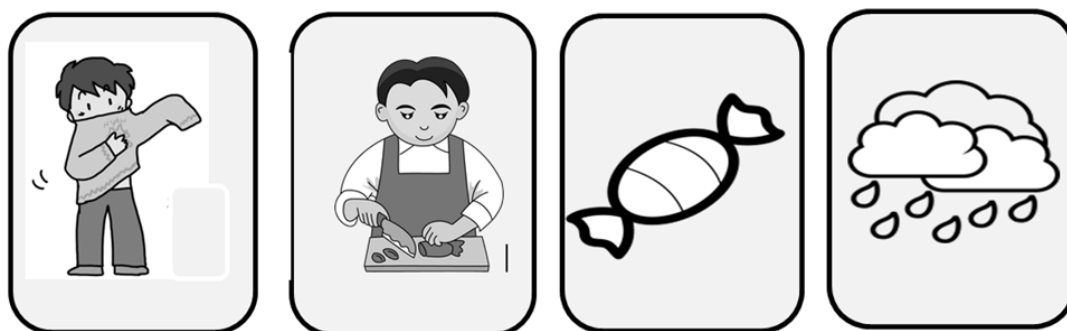
Level: Beginner

Step: Awareness, listening discrimination, controlled practice, guided practice

Also: *arimasu/arimasen*

Activity: Go Fish

The instructor creates sets of cards with the pictures of pitch-accent minimal pairs (to wear/to cut, rain/candy, etc.), as shown in the illustration below. Several small groups receive one set and pass out all the cards to the members of that group. Each student must in turn select another student and ask them if they have the matching card: If one student has the 'rain' card (*ame*), then they need to ask another student for another 'rain' card. They could use the grammar point *x ga arimasu ka* 'Do you have x?' to ask for the card they want. The other student would answer *hai, arimasu. hai, doozo* 'Yes, I have (it). Here you go.' if they have the 'rain' card or *iie, arimasen* 'No, I do not have (it).' Minimal pairs differing by vowel/consonant length or words with problematic segmentals could be included.



Topic: Loanwords

Level: Beginner/Intermediate

Step: Awareness, listening discrimination, communicative practice

Also: *katakana*, long vowels/consonants, ordering food, making a commercial, role-play, polite language, pragmatics, numbers/money, etc.

Activity: Ordering fast food

Loanwords are ideal for working on pitch accent as they are 'known words' (i.e., low cognitive load). Instructors can quickly tell the students the pattern for pitch accent placement or have students listen and guess the pattern. Practice can be embedded into lessons on ordering fast food (*-o kudasai* 'please give me x') or practically any grammar point, e.g., *McDonald's shakes are more delicious than Burger King's shakes* (comparatives). Students could also make a commercial advertising product.

Topic: 20 Questions/I spy game

Level: Beginner/Intermediate

Step: Listening discrimination, controlled practice, guided practice

Also: *hiragana*, modifying nouns

Activity: Asking questions 1

The instructor prepares objects, picture cards or handouts with pictures. The instructor models the activity having the students ask yes/no questions such as ‘Is it rain?’ (minimal pairs) or ‘Is it a red car?’ (adjective + noun), to guess what the instructor is thinking of. Afterwards, the students divide into smaller groups with one student imagining the object/picture and the others guessing. Also, the instructor could write a word/draw a picture on a piece of paper and attach a different one to each student’s forehead and have them mingle and ask each other questions about what is written on their forehead.

Topic: Describing habits, activities, etc.

Level: Intermediate

Step: Guided practice, communicative practice

Also: Verb forms, pragmatics

Activity: Asking questions 2

When reviewing or introducing new verb forms, the instructor can ‘mix-up’ the forms to practice shifting pitch accent patterns. For example, the instructor asks students what they or someone else did on the weekend/last night/last summer/etc. and then, what they or someone else does every weekend/every night/every summer/etc. The instructor can add other forms, e.g., *-masu* form or even the humble/honorific forms to practice shifting pitch accent patterns (and grammatical forms, pragmatics).

Topic: Academic talks

Level: Intermediate/Advanced

Step: Awareness, listening discrimination, controlled practice

Also: Grammar, vocabulary, rhetoric, speaking fluency, other pronunciation features

Activity: Recreating a talk

Students listen to a talk or excerpt (e.g., TEDxTokyo). The instructor has students mark a transcript for thought groups (where they should pause), fall in pitch, pitch patterns, etc. The students shadow along with the video, soundfile or instructor. Students also orally present the talk using the transcript or an outline. Students record themselves and self-evaluate their recording and/or receive feedback from the instructor.

Topic: Kansai dialect/folktales

Level: Intermediate/Advanced

Step: Awareness, listening discrimination

Also: Grammar, vocabulary, other pronunciation features

Activity: Noticing dialectal features

Instructors have students listen to sample sentences, dialogs or talks, using websites, TV shows, or movies, and make comparisons with standard Japanese. Some useful websites for the Kansai dialect include: <http://www.kansai-ben.com/> and http://www.eastudies.org/guide_kansai-ben.html and for folktales: <http://minwa.fujipan.co.jp/hagukumu/minwa/>. The students are put into small groups to point out the differences in pronunciation (and vocabulary, grammar). The instructor explains the features or has students compare dialogs in standard Japanese and Kansai Japanese. For instructors proficient in the Kansai dialect, they could create entire activities/lessons in a communicative framework, using the drawings and exercises from a textbook lesson for standard

Japanese replaced with the Kansai dialect. For example, the lessons on *shinakya* 'have to do' in standard Japanese could be easily modified for *sena akan* 'have to do' in the Kansai dialect.

For the folktales, the instructor could have the students listen to the folktale and then, orally recreate the story they heard in standard Japanese. Students could point out the differences they noticed with the instructor adding explanations.

CONCLUSION

Pitch accent is an essential part of mastering not only Japanese pronunciation, but also global proficiency due to its connections to grammar, vocabulary, expository speech, reading, and pragmatics. As such, it is vital that time and effort be devoted to aiding learners in acquiring pitch accent. We hope these teaching tips will open up dialog among teachers concerning the teaching of pitch accent in JFL/JSL classes and offer suggestions as to how to do so effectively within a communicative framework.

ACKNOWLEDGEMENTS

We thank Chisato Kojima for her comments.

ABOUT THE AUTHORS

Vance Schaefer (PhD, Second Language Studies, Indiana University, schaefer@olemiss.edu) is an assistant professor of TESL and applied linguistics in the Department of Modern Languages at the University of Mississippi. He is an experienced JFL and EFL/ESL instructor. His research interests include L2 phonology, specifically lexical pitch perception and foreign accent, and pronunciation pedagogy.

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

REFERENCES

- Baker, W., & Smith, L. C. (2010). The impact of L2 dialect on learning French vowels: Native English speakers learning Québécois and European French. *The Canadian Modern Language Review / La revue canadienne des langues vivantes*, 66, 711-738.
- Celce-Murcia, M., Brinton, D., Goodwin, J., & Griner, B. (2010). *Teaching pronunciation: A course book and reference book* (2nd Edition). Cambridge: Cambridge University Press.

- Hirano-Cook, E. (2011). *Japanese pitch accent acquisition by learners of Japanese: Effects of training on Japanese accent instruction, perception, and production* (Doctoral dissertation). Retrieved from KU ScholarWorks. (11708)
- Gatbonton, E., & Segalowitz, N. (1988). Creative automatization: Principles for promoting fluency within a communicative framework. *TESOL Quarterly*, 22, 473–492.
- Kawahara, S. (in press, 2015). The phonology of Japanese accent. In H. Kubozono (Ed.), *The handbook of Japanese language and linguistics: Phonetics and phonology*. Berlin: Mouton.
- Kawasome, Y. (2014, April). hatsuon gakushuu no furikaeri ga nihongo kyoushi no onsei kyouikukan ni ataeru eikyou – kyouikusha to gakushuusha no futatsu no tachiba de no katari kara [The influence of (students') reflecting on learning pronunciation on Japanese instructors' pronunciation teaching – accounts from the perspectives of both educators and learners]. Poster presented at research meeting of Tokyo Association of Phonetics and Speech (TAPS), Tokyo, Japan.
- Kindaichi, H. (1988). *The Japanese language*. Tokyo: Tuttle Publishing.
- Kondo, M. (2007). Acoustic realization of lexical accent and its effects on phrase intonation in English speakers' Japanese. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 1649-1652). Dudweiler: Pirrot.
- Kubozono, H. (2012). Varieties of pitch accent systems in Japanese. *Lingua*, 122, 1395-1414.
- Loschky, L., & Bley-Vroman, R. (1993). Grammar and task-based methodology. In G. Crookes & S. Gass (Eds.), *Tasks and language learning: Integrating theory and practice* (pp. 123-167). Clevedon, England: Multilingual Matters Ltd.
- Otake, T., & Cutler, A. (1999). Perception of suprasegmental structure in a nonnative dialect. *Journal of Phonetics*, 27, 229-253.
- Rvachew, S., Nowak, M., & Cloutier, G. (2004). Effect of phonemic perception training on the speech production and phonological awareness skills of children with expressive phonological delay. *American Journal of Speech-Language Pathology*, 13, 250-263.
- Sato, T. (1995). A comparison of phonemes and prosody in the evaluation of spoken Japanese. *Japanese Language Education Around the Globe*, 5, 139-154.
- Sato, Y., Utsugi, A., Yamane, N., Koizumi, M., & Mazuka, R. (2013). Dialectal differences in hemispheric specialization for Japanese lexical pitch accent. *Brain and Language*, 127, 475-483.
- Segalowitz, N., & Hulstijn, J. (2005). Automaticity in second language learning. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 371–388). Oxford: Oxford University Press.
- Shibata, T., & Shibata, R. (1990). Akusento wa doongo o dono teido benbetsu shiuru ka – nihongo, eigo, chuugoku no baai [To what extent can accent differentiate homophones? The case of Japanese, English and Mandarin]. *Keiryoo Kokugaku*, 17, 311-323.

- Shport, I. A. (2008). Acquisition of Japanese pitch accent by American learners. In P. Heinrich and Y. Sugita (Eds.), *Japanese as foreign language in the age of globalization* (pp. 165-187). German Institute for Japanese Studies. München: Iudicium Verlag.
- Sicola, L., & Darcy, I. (2015). Integrating pronunciation into the language classroom. In M. Reed & J. Levis (Eds.), *The Handbook of English pronunciation* (pp. 471-487). Malden, MA: Wiley Blackwell.
- Sugiyama, Y. (2012). *The production and perception of Japanese pitch accent*. Cambridge: Cambridge Scholars Publishing.
- Tanaka, S., & Kubozono, H. (2012). *Introduction to Japanese pronunciation: Theory and practice*. Tokyo: Kurosio.
- Tanimori, M., & Sato, E. (2012). *Essential Japanese grammar: A Comprehensive guide to contemporary usage*. North Clarendon, VT: Tuttle Publishing.
- Tsurutani, C. (2011). L2 Intonation: A study of L1 transfer in Japanese intonation by English-speaking learners, *Second Language*, 10, 79-102.
- Wang, Y., Jongman, A., & Sereno, J. (2003). Acoustic and perceptual evaluation of Mandarin tone production before and after perceptual training. *Journal of the Acoustical Society of America*, 113, 1033-1044.